



XI

JORNADAS DE  
COOPERACIÓN  
BIBLIOTECARIA

**GRUPO DE TRABAJO DE PATRIMONIO DIGITAL**

**Jornadas de Cooperación Bibliotecaria, 2007**  
**Ministerio de Cultura y Comunidades Autónomas**  
Mérida, 20, 21 y 22 de Marzo

COORDINADORES: Teresa Malo de Molina.  
Directora Técnica de la Biblioteca Nacional. Madrid  
Alejandro Carrión Gútiez  
Director de la Biblioteca de Castilla y León. Valladolid



## GRUPO DE TRABAJO DE PATRIMONIO DIGITAL INFORME

### 1. Antecedentes

La 8ª conclusión de las X Jornadas de Cooperación, celebradas en Santiago de Compostela, se refería las líneas de actividad del grupo de trabajo de Patrimonio Digital:

“El grupo de trabajo de *Patrimonio digital* trabajará en la puesta en marcha de las conclusiones de las Jornadas de Preservación de Patrimonio Digital, especialmente en la creación de la “Red de Conocimiento” propuesta en dichas conclusiones, así como en la difusión y selección de los contenidos relacionados con la materia y recogida de legislación sobre este tipo de materiales.”

La actividad del grupo en estos meses se ha centrado en el seguimiento de los proyectos que se encuentran en fase de estudio o de ejecución para comunicar la experiencia derivada de ellos a los organismos interesados en llevar a cabo iniciativas relacionadas con la conservación del patrimonio digital.

### 2. Miembros del grupo

Durante el presente año no ha habido cambios en la composición del grupo de trabajo. Forman parte del Grupo de Trabajo de Patrimonio Digital los siguientes miembros en representación de las Comunidades Autónomas y el Ministerio de Cultura:

- Laura Cerezo Navarro. Biblioteca de Andalucía. Granada
- Isabel Móneva. Biblioteca de Aragón. Zaragoza
- Santiago Caravia Nogueras. Biblioteca de Asturias. Oviedo
- José María Gutiérrez Rodríguez. Biblioteca Central de Cantabria. Santander
- Castilla-La Mancha. Pendiente de designar representante.
- Eugènia Serra Aranda. Biblioteca de Catalunya. Barcelona
- Ana Leria Ayora. Biblioteca Pública de Ceuta. Ceuta
- José Luis González González. Centro Superior Bibliográfico de Galicia. Santiago de Compostela.
- Javier Muñoz Aragón. Servicio Regional de Bibliotecas y del Libro. Comunidad de Madrid. Madrid.
- Fina Sánchez Fernández. Biblioteca Regional de Murcia. Murcia
- Juan Francisco Elizari Huarte. Biblioteca General de Navarra. Pamplona
- Francisca Pulgar Vernalte. Servicio de Bibliotecas. Gobierno Vasco. Vitoria
- Romà Seguí Francés. Biblioteca Valenciana. Valencia
- Isabel García-Monge Carretero. Subdir. Gral. Coordinación Bibliotecaria. Ministerio de Cultura.



Coordinadores:

- Teresa Malo de Molina. Biblioteca Nacional. Madrid
- Alejandro Carrión Gútierez. Biblioteca de Castilla y León. Valladolid

### **3. Objetivos**

Como se ha comentado, el objetivo principal del trabajo del grupo durante este año ha sido el estudio de los proyectos que en la actualidad se encuentran en fase de estudio y, sobre todo, en producción, ya que la experiencia práctica derivada del desarrollo de las distintas fases de los proyectos es de gran interés para orientar proyectos similares y para construir una base de conocimiento relacionada con los distintos aspectos de la preservación del patrimonio digital.

### **4. Reunión del Grupo de trabajo**

El Grupo de trabajo de Patrimonio Digital ha celebrado una reunión el día 7 de marzo de 2007 en la Biblioteca Nacional, Madrid

La reunión se celebró con el siguiente orden del día:

- Proyecto PADICAT
- Convocatoria de subvenciones del Ministerio de Cultura para proyectos de digitalización y preservación de contenidos digitales
- Otras iniciativas relacionadas con el patrimonio digital
- Propuestas para una política de preservación del patrimonio digital

### **5. Proyecto PADICAT de la Biblioteca de Catalunya**

PADICAT es el primer proyecto de archivo web operativo en el Estado Español. Está coordinado por la Biblioteca de Catalunya. El Centre de Supercomputació de Catalunya (CESCA) proporciona el soporte y la infraestructura informática al proyecto.

En documento Anexo al presente informe se incluye información detallada del proyecto PADICAT elaborada por Eugènia Serra, Responsable de la Coordinación General de la Biblioteca de Catalunya.

### **6. Convocatoria de subvenciones del Ministerio de Cultura**

El Ministerio de Cultura tiene previsto publicar en breve una Convocatoria de subvenciones para proyectos de digitalización y preservación de contenidos digitales. La representante de la Subdirección General de Coordinación Bibliotecaria del Ministerio de Cultura en el grupo de trabajo, Isabel García-Monge Carretero, informó sobre las características generales de la convocatoria. De acuerdo con lo informado, existirán dos líneas de subvención: una para administraciones públicas, dotada con



1.650.000 euros, y otra para entidades sin ánimo de lucro, con una cuantía total de 300.000 euros.

Las ayudas se podrán justificar hasta marzo de 2008 y no podrán destinarse a equipamiento informático.

### **7. Otras iniciativas relacionadas con la preservación digital**

Teresa Malo de Molina, Directora Técnica de la Biblioteca Nacional, informa sobre las iniciativas de la BNE en relación con la conservación del patrimonio digital:

- Durante el año 2007 se desarrollará un proyecto de puesta en marcha de un servidor de DOIs (Digital Object Identifier) para ofrecer a los productores de contenidos digitales en España la posibilidad de utilizar este servicio y plantear la entrega de contenidos para su preservación la BNE.
- A partir de 2008 se desarrollará el proyecto de Archivo de Internet. En este momento la BNE a la búsqueda de un socio tecnológico.

### **8. Propuestas para establecer una política estatal de preservación digital**

Se insiste en la necesidad de elaborar unas directrices que puedan orientar una estrategia nacional en materia de preservación digital,

En coordinación con el grupo de trabajo de DL se considera prioritario insistir en la necesidad de promulgar una legislación estatal en materia de Depósito Legal que facilite la conservación del patrimonio digital.

Madrid, marzo de 2007

## ANEXO

### ESTADO DE DESARROLLO DE PADICAT (Patrimonio digital de Catalunya) ([www.padicat.cat](http://www.padicat.cat))

#### DATOS GENERALES

Coordinación y definición del proyecto: Biblioteca de Catalunya, ([www.bnc.cat](http://www.bnc.cat))  
Soporte e infraestructura informática: Centre de Supercomputació de Catalunya, (CESCA, [www.cesca.cat](http://www.cesca.cat))

Ambas entidades son organismos autónomos vinculados a la Generalitat de Catalunya. El CESCA es un consorcio público vinculado al Departamento de Educación, Universidades e Innovación y la Biblioteca de Catalunya es un organismo autónomo del Departamento de Cultura.

Ambos tienen capacidad jurídica para establecer convenios que firman sus respectivos directores.

Presupuesto plurianual 2006-2008: 766.000 € (que gestiona la BC)

Equipo técnico dedicado al 100% al proyecto: 1 coordinador + 1 administrativo + 1 documentalista + 1 informático. En breve se prevé la incorporación de dos documentalistas más.

#### LINEAS DE TRABAJO INICIADAS

**Convenios con productores de patrimonio digital de Catalunya: instituciones culturales y académicas, sociedad civil, administración pública, medios de comunicación, empresas, ...**

- Mediante el convenio autorizan la captura, preservación y difusión de los contenidos de sus Webs.
- Algunos datos:
  - Entidades identificadas, es decir con datos fiables de contacto: 1076
  - Cartas enviadas, firmadas por la directora de la BC: 848
  - Convenios formalizados: 105
- Se ha elaborado un convenio modelo (que se adjunta). En el caso de instituciones con contenidos en sus webs de los que no ostentan los derechos de autor se introduce una fórmula por la cual la BC se responsabiliza de gestionar la petición de estos de derechos.

#### **Acontecimientos relevantes para la historia de Catalunya**

##### Elecciones al Parlament de Catalunya

Se realizó una captura intensiva de webs relacionadas con las elecciones autonómicas de noviembre 2006.



## JORNADAS DE COOPERACIÓN BIBLIOTECARIA

- Se capturaron con frecuencia alta (diaria, 2 veces por semana o semanalmente, dependiendo de cada caso), webs de los partidos políticos con o sin representación parlamentaria, blogs de los candidatos, webs de fundaciones de carácter político, y medios de comunicación digitales, creadores de opinión, ...
- Se realizaron las capturas desde 15 días antes de las elecciones hasta 15 días después aproximadamente.
- Un total de 652 capturas de 83 webs diferentes.
- Esta información se presenta agrupada con espacio propio en el portal PADICAT

### Captura de otros dominios

- Se ha firmado un acuerdo con la Fundación PuntCAT por el cual la Fundación facilitará a la Biblioteca de Catalunya la relación de dominios .CAT registrados, para su captura. Son aproximadamente 19.000 dominios
- Se recogen propuestas de incorporación de Webs mediante un formulario en el portal PADICAT. Estas propuestas se validan antes de incorporarlas al proyecto
- Algunos de los dominios son coincidentes con los ya identificados por convenios.

## HARDWARE Y SOFTWARE

Marzo-junio 2006 → instalación del hardware y del software en el Centre de Supercomputació de Catalunya

Julio de 2006 → inicio de las capturas a partir de las líneas de trabajo expuestas anteriormente.

Una vez realizadas las primeras capturas, especialmente las de las elecciones al Parlament, se ha podido hacer una primera valoración del funcionamiento del hardware y del software.

### A destacar que:

- De los procesos vinculados a PADICAT (captura, almacenamiento, indización y recuperación), el proceso que más recursos del sistema precisa, con diferencia, es la indización de la información.
- La eficacia del sistema va muy vinculada a la forma de realizar y programar los diferentes procesos (trabajando por lotes o realizando todos los procesos simultáneamente).
- Existe un software que no conocíamos a priori que permite validar los recursos capturados y asignar metadatos, es el *Web Curator*. A partir del mes de marzo se utilizará este software para ambas funciones. En relación con los metadatos se pretende asignarlos a aquellos recursos que se consideren de interés especial, de manera que al recuperar la información aparezcan en primer lugar en la lista de resultados.
- Hay dos interfaces diferentes que pueden trabajar con Heritrix que es el que se utiliza en la mayoría de proyectos; una es Wera (es la que actualmente utiliza la BC) y la otra es WayBack (la de Internet archive). Tanto una como otra presentan algunas deficiencias o limitaciones a causa de que no se han desarrollado demasiado. Desde la BC/CESCA se analizarán los próximos meses y se intentará desarrollarlos para sumar las opciones que ofrecen ambos, u ofrecer el resultado a la comunidad que utiliza Heritrix.



## ACCIONES A CORTO PLAZO

### Equipo técnico

- Incorporación de dos documentalistas más en los próximos meses
- Incorporación de un analista programador para desarrollar la interfaz de consulta

### Hardware

- Ampliación del número de procesadores par mejorar la capacidad de indización (a partir del mes de abril)

### Capturas intensivas

- Programación de capturas intensivas con motivo de las elecciones municipales en Catalunya:
  - Webs de partidos políticos, candidatos y medios de comunicación en capitales de comarca y segunda población de mayor tamaño (básicamente serán municipios con más de 20.000 hab., más una muestra representativa del territorio de municipios menores (menos de 5.000 hab.)
  - Capturas entre abril y junio
  - Diferentes frecuencias dependiendo de cada caso.

## CIFRAS DEL PROYECTO A FEBRERO 2007

Número de webs diferentes: 250

Número de captures: 800

Número de archivos: 4,5 millones

Espacio en disco: 230 GB

Entidades identificadas para firmar convenio: 1076

Entidades contactadas: 848

Convenios firmados: 105

Evaluación del portal web del sistema (cómputo estadístico iniciado en diciembre de 2006)

Visites a la web (diciembre-febrero): 2989

Visitantes diferentes (diciembre-febrero): 2096

Difusión del proyecto en entornos profesionales

Artículos científicos en revistas o congresos: 5

Actos de presentación del proyecto a público profesional: 15

Difusión genérica del proyecto en los medios de comunicación

Notas de prensa: 7

Presencia en los medios de comunicación: 28

Eugènia Serra

Responsable de la Coordinación General

Marzo, 2007