

MEI
II, vol. 2
n° 2

La Biblioteca de Catalunya i l'accés al patrimoni digital

Eugènia Serra, Karibel Pérez y Ciro Llueca
Biblioteca de Catalunya

Resumen

Desde 2004 la *Biblioteca de Catalunya* (BC) se ha posicionado para abordar el reto del nuevo paradigma tecnológico bajo la misión de convertirse en una biblioteca "abierta, fiable y útil". Con la vista puesta en la próxima modificación de la ley del depósito legal, la estrategia de difusión del patrimonio digital contempla las acciones destinadas a publicar en Internet las colecciones analógicas de la BC, y complementariamente la creación del archivo de las páginas web de Cataluña. El último paso en la consolidación de esta nueva política bibliotecaria es un instrumento de preservación en alta seguridad creado a partir de las recomendaciones e iniciativas internacionales y de la experiencia de la propia BC.

Recibido el
06-05-2011

Aceptado en
25-07-2011

Palabras clave

Bibliotecas nacionales; Preservación digital; Repositorios digitales.

Abstract

Since 2004 the *Library of Catalonia* (BC) has positioned itself to address the challenge of the new technological paradigm under the mission of becoming a library "open, reliable and useful". With a view to the forthcoming amendment of the law of legal deposit, the strategy of digital heritage promotion includes actions to publish online the analog collections of BC, and the creation of the Web archive of Catalonia. The last step in the consolidation of this new library policy is a high security preserving tool, based on the international guidelines and initiatives and the experience of the BC itself.

Keywords

National libraries; Digital preservation; Digital repositories.

La Biblioteca de Catalunya en l'escenari digital

Des de fa uns anys les biblioteques, arxius i museus ens trobem immersos en un cicle permanent de creació, identificació, captura, accés i preservació de documents i de recursos d'informació com probablement no s'havia vist abans. La tendència està consolidada.

Les tecnologies de la informació han proporcionat vies de fer accessibles documents que no havíem ni imaginat: cançoners provençals, manuscrits, fotografies o mapes antics estan ara a l'abast de tothom gràcies a la digitalització i a Internet. Al mateix temps, la web ha proporcionat a les institucions, les empreses i els particulars noves formes de comunicar, informar, formar i interactuar amb la societat, els usuaris i altres organitzacions. I finalment, el món de les indústries culturals, els editors de llibres i publicacions o els creadors de música i de pel·lícules, han

iniciat un canvi de model de negoci obligats per la mateixa evolució de les tecnologies, per l'actual context econòmic, i pels canvis d'hàbits i forma de relacionar-se de les persones.

En aquest context, les institucions de la memòria han diversificat el patrimoni a difondre i preservar quant al seu tipus, forma de creació i vies d'adquisició. Això implica necessàriament repensar la forma de gestionar, els processos, els recursos i equipaments, la tecnologia i les capacitats dels professionals.

L'escenari es consolidarà properament si el nou projecte de *Llei del Dipòsit Legal* (DL), aprovat pel Consell de Ministres, completa la tramitació al Congrés i esdevé una ja urgent nova llei que reguli la recollida i preservació dels documents nascuts digitals i que habiliti a les institucions patrimonials a aplicar les tècniques de transformació i tractament necessàries per garantir-ne la pervivència¹.

El projecte de llei presenta tres novetats rellevants: la primera és que l'obligació del DL recau en l'editor o productor, promotor real de la creació del document; superant el model de producció vigent durant segles, en que l'impressor era a la vegada editor.

La segona novetat és que reconeix la necessitat de preservar la producció digital en suport i en línia, i la competència de les comunitats autònomes de capturar, guardar i transformar aquests documents.

La tercera novetat, en aplicació de les competències en cultura transferides des de fa anys a les autonomies, reconeix el criteri de les comunitats per decidir com i què recollir, tan a efectes de patrimoni digital, com, en part, als documents tradicionals.

L'aprovació de la nova llei suposarà un increment en el volum de llibres i revistes electròniques que arribaran per dipòsit legal a les biblioteques dipositàries (anomenades "centres de conservació" al nou projecte de llei)².

En el pla estratègic de l'any 2004³ i les seves successives edicions, la *Biblioteca de Catalunya* (BC) s'ha posicionat activament per abordar aquesta nova realitat i aprofitar-ne al màxim les oportunitats. Per més que soni a tòpic o a utopia, la biblioteca sense parets ni limitacions físiques està més a prop que mai, i paral·lelament la visió de la BC de ser "oberta, fiable i útil" sembla més assequible.

Així, l'any 2005 va suposar un punt d'inflexió al respecte per al sistema bibliotecari de Catalunya, i especialment per la BC⁴, que basava en tres eixos la seva política de difusió i preservació del patrimoni digital: en primer lloc, la posada en marxa de projectes per digitalitzar retrospectivament les seves col·leccions analògiques, com la MDC (*Memòria Digital de Catalunya*), ARCA (*Arxiu de Revistes Catalanes Antiques*) i el projecte *Google Llibres*.⁵

En segon lloc, el disseny, creació i producció de PADICAT (*Patrimoni Digital de Catalunya*), un repositori per al patrimoni nascut digital: les pàgines web publicades a Internet. En tercer lloc, el disseny i creació de COFRE (*CO*nservem per al *F*utur

Recursos Electrònics), un instrument de preservació en alta seguretat creat a partir de la pròpia experiència.

Vegem a continuació una explicació acurada dels tres protagonistes de la política d'accés al patrimoni digital de la BC.

Digitalització retrospectiva a la *Biblioteca de Catalunya*

Una de les funcions comunes de les biblioteques nacionals arreu del món és la de preservació del patrimoni. Dècades enrere la preservació es basava en accions de conservació sobre els objectes físics per mantenir-los en el millor estat possible, el seu emmagatzematge en les condicions adequades i la seva microfilmació, que permetia disposar d'una còpia màster fidel a l'original.

Actualment, la microfilmació ha estat substituïda en bona part de les institucions per la digitalització, que proporciona substituïts digitals d'alta qualitat i dona el *plus* de poder-se difondre fàcilment per les xarxes de comunicació, facilitant així el descobriment de documents fins aleshores quasi "amagats" a la comunitat.

La digitalització de fons patrimonials ajuda a preservar els originals analògics perquè disminueix la manipulació dels originals, ja que bona part de les necessitats d'ús es resolen amb l'accés a la còpia digital; i a la vegada genera objectes digitals que cal preservar per evitar noves manipulacions dels originals i per disposar de còpies fidels permanents.

Atesos els costos assequibles de la digitalització i la publicació a Internet dels fons i col·leccions documentals, l'experiència adquirida amb els anys en l'ús d'aquesta tecnologia per part de les biblioteques i empreses de serveis, i el rendiment i valor que proporciona en termes de servei als usuaris sens limitacions de temps i espai, el volum de fons patrimonials digitalitzats s'incrementa any rere any.

El rendiment d'aquestes iniciatives de digitalització en termes d'accés i difusió queda ben palès si atenem a les estadístiques d'ús dels documents digitalitzats; a mode d'exemple, els 275 títols complets de revistes i diaris que ofereix ARCA generen al voltant de 130.000 consultes mensuals. Es tracta doncs d'un ús intensiu que reforça l'estratègia de la BC d'invertir recursos en continuar aquesta línia d'actuació.



Fig. 1. ARCA (*Arxiu de Revistes Catalanes Antiques*)

La BC ha focalitzat la seva política de digitalització en documents en situació de domini públic, que no tenen drets d'autor vigents; i d'acord amb els següents criteris de selecció: documents fràgils, documents singulars, documents amb interès per a la recerca, documents que presentin un estat físic adequat per a la manipulació i digitalització, documents que tinguin requeriments d'ús per a la consulta dels originals (p.e. les plaques de vidre fotogràfiques requereixen d'una taula de llum), documents que estiguin complets (especialment les publicacions en sèrie) i documents que siguin de consulta freqüent.

Complementàriament, la BC ha volgut traduir les polítiques a repositoris, tot treballant cooperativament amb altres centres. MDC (*Memòria Digital de Catalunya*)⁶ és una de les iniciatives de treball cooperatiu que lidera la BC conjuntament amb el *Consorci de Biblioteques Universitàries de Catalunya*. En aquest repositori cooperatiu la BC hi té creades 14 col·leccions digitals de manuscrits, incunables, materials gràfics, fotografies, fons d'arxiu, mapes, materials gràfics, partitures, enquadernacions artístiques i fons personals, que suposen més de 20.000 documents que representen un total aproximat de 200.000 imatges. MDC dona accés a un total de 47 col·leccions de caràcter patrimonial procedents de 17 institucions culturals i de recerca.

ARCA (*Arxiu de Revistes Catalanes Antiques*)⁷, obert a Internet el maig del 2005, és un altre projecte cooperatiu que posa a l'abast dels usuaris 275 títols complets de publicacions periòdiques representatives de la cultura i la societat catalanes, que ja no es publiquen. Són capçaleres d'informació general, art, literatura, llengua, medicina, ciències, pedagogia, política, història i humorisme gràfic que s'han completat gràcies a la participació de més de 21 biblioteques, arxius, museus, ateneus i associacions. ARCA suma més de 600.000 pàgines digitalitzades que aquest mateix 2011 arribaran al milió dues-centes mil pàgines amb la incorporació de títols tan emblemàtics com *La Veu de Catalunya* o *El Diluvio*.

El gener del 2007 la BC anuncià l'acord amb *Google* per participar en el projecte *Google Llibres*, i liderar la incorporació dels fons no només de la BC sinó de altres quatre biblioteques patrimonials catalanes. Aquest acord ha permès a la BC digitalitzar amb un cost reduït al voltant de 60.000 documents que es troben accessibles i disponibles per ser descarregats a la plataforma de *Google* i que representen prop d'un milió d'imatges; de tots aquests documents la BC n'obté una còpia que emmagatzema com a còpia de preservació.

Tots els objectes digitals creats en el marc de les iniciatives i projectes esmentats són ara patrimoni digital que cal preservar. A la BC entenem per patrimoni digital a preservar els objectes digitals fruit de la transformació d'originals analògics que ha dut a terme la mateixa biblioteca individual o cooperativament; els recursos digitals en suports portàtils (CD i DVD); i els recursos nascuts digitals publicats a Internet, com són les pàgines webs, els blogs, els llibres, diaris i revistes electròniques, etc.

En relació amb l'arxiu de la web catalana, la BC, emmirallant-se en biblioteques nacionals d'arreu del món, va crear un singular repositori, el PADICAT.

L'arxiu d'Internet a la Biblioteca de Catalunya

L'arxiu d'Internet és actualment una realitat arreu del món tecnològicament desenvolupat, tot i la imperfecció de la major part dels sistemes informàtics que serveixen a les polítiques nacionals de preservació del patrimoni digital en la xarxa. Creat el 2005, el PADICAT (*Patrimoni Digital de Catalunya*) és un repositori destinat a compilar, processar i donar accés permanent a la producció digital catalana a Internet. És, sintèticament, l'arxiu web de Catalunya, dedicat a preservar els recursos digitals, essencialment pàgines web, publicats a Internet per al públic de Catalunya.

A partir d'una fase inicial d'anàlisi dels dipòsits existents⁸, el dipòsit ha experimentat un període de naixement (2005-2006), creixement (2007-2008) i consolidació (2009-2011), que és vigent en el moment de redacció d'aquest article.

PADICAT no pretén ni podria capturar sistemàticament totes i cadascuna de les versions canviants de les webs catalanes, però sí té com a propòsit proporcionar una fotografia representativa i significativa de la web catalana capturada al menys dos cops a l'any, que perduri amb el pas del temps.

Des de l'11 de setembre del 2006 la BC manté operatiu i actualitzat el portal <<http://www.padicat.cat>>, en català, castellà i anglès. Tota la col·lecció és accessible en obert i en línia, consultable per cerca, per navegació a directori temàtic, o per accés directe a paquets monogràfics.

D'acord amb el model híbrid, tendència generalitzada en repositoris similars d'arreu del món, la política de col·lecció del dipòsit es basa en les accions següents:

- Compilar massivament els recursos digitals publicats en obert a Internet, per mitjà de la captura del domini .cat.
- Impulsar el dipòsit sistemàtic de la producció web de les entitats i les empreses de Catalunya, mitjançant la identificació i la signatura d'un conveni de cooperació.
- Promoure línies de recerca per mitjà de la presentació temàtica dels recursos digitals capturats relatius a determinats esdeveniments de la vida pública catalana, com ara campanyes electorals a Internet, el fenomen de la música en línia, o els museus a Internet⁹.

Després de cinc anys d'existència, el dipòsit conté 118.587 captures de 39.587 pàgines web i està format per 250 milions de fitxers informàtics, amb una mida de 7,7 TB¹⁰; i la seva previsió de creixement anual és de 75.700 versions, d'aproximadament 32.000 pàgines web.

Pel que fa als aspectes legals, des del plantejament inicial del dipòsit, les limitacions jurídiques han estat analitzades amb rigor però també amb lògica. Més enllà dels condicionants legals, la BC ha compartit la defensa d'una filosofia explotada amb èxit des del 1996 per l'*Internet Archive*, segons la qual la captura de les parts públiques d'Internet és bàsica per preservar la cultura i el patrimoni de la nostra comunitat, igual que han fet les biblioteques amb els llibres, les revistes, els discos i les pel·lícules que al llarg del temps hi ha hagut. Com ha indicat Josep Vives (2005)¹¹, les administracions i els professionals disposem d'arguments bons i suficients per convèncer els nostres dipositants de la bondat dels repositoris, sense entrar en debats estèrils sobre la legalitat o no de preservar la producció digital.

Dos aspectes que creiem ha contribuït a vèncer qualsevol resistència respecte a l'accés públic que es fa del contingut de PADICAT és per un costat, la forma de presentació dels continguts (en forma cronològica) o organitzats temàticament, i per l'altra, el fet que es bloquegi l'indexació dels continguts per part dels motors de cerca; aquestes dues accions signifiquen clarament la voluntat de la BC de no "competir" amb les webs oficials i actualitzades, i redueixen les pors per part dels creadors de les webs a que siguin capturades.

Partint d'aquesta seguretat i en compliment de la política de col·lecció basada en els agents productors de les pàgines web a Catalunya, la BC ha signat 450 convenis de cooperació amb entitats i empreses de tots els sectors, que formalment li permeten capturar, processar i preservar les captures fetes dels seus recursos digitals i donar-hi accés obert.

Pel que fa a l'arquitectura tècnica del sistema, que compra amb l'activa col·laboració del CESCA (*Centre de Supercomputació de Catalunya*), posteriorment a la fase d'anàlisi i test de programari es va determinar que s'utilitzaria el programa informàtic *Heritrix*¹², emprat en la major part de projectes de captura de recursos digitals. Aquest és el programa encarregat de compilar les pàgines web tal com les veu l'usuari que navega per Internet i emmagatzemar-les en arxius comprimits en format ARC¹³.

A continuació, el programari *Heritrix* es complementa amb *NutchWax*¹⁴, o bé la combinació d'*Hadoop*¹⁵ i *Wayback*¹⁶, que duen a terme uns processos d'indexació de la informació compilada que permeten, ulteriorment, utilitzar aquests índexs per localitzar els recursos dins de la col·lecció mitjançant les seves respectives interfícies de consulta: *Wera*¹⁷, que permet la cerca per paraules clau a través dels índexs generats per *NutchWax*; i *Wayback*, que permet la consulta directa per URL en els índexs generats per *Hadoop* i el mateix *Wayback*.

Finalment, s'ha aprofitat el programa *Web Curator Tool*¹⁸, desenvolupat per la *National Library of New Zealand* i la *British Library*, com a sistema de gestió documental que permet l'assignació de metadades a una part significativa de la col·lecció, amb la intenció de poder integrar, en el futur, els fons del dipòsit a la cerca en altres catàlegs, tant de la *Biblioteca de Catalunya* com d'altres institucions.

D'altra banda, el personal del CESCO, soci tecnològic del projecte, ha desenvolupat i compartit amb la comunitat diverses aplicacions *ad hoc*, com els mòduls del CAT (*Curator Archiving Tool*), dissenyats per millorar l'accés i la recuperació dels recursos digitals dipositats al PADICAT¹⁹. Tot el programari emprat és de codi obert i gratuït.

Pel que fa al maquinari que sosté el sistema, es compta amb sis nodes *HP ProLiant DL360 G4p*, encarregats de les tasques de recol·lecció i indexació de les pàgines web. De la cerca i la visualització de resultats en la interfície web, se n'encarrega un clúster *Linux* d'alta disponibilitat amb característiques de balanceig de càrrega de peticions i de tolerància d'errors en cas de desastre tècnic dels nodes que integren la plataforma. Una cabina *NetApp FAS3170* presenta un espai de disc via NFS a aquests nodes. El sistema es completa amb un robot on es conserven còpies de seguretat de les dades en cinta.

A començament del 2006 es van fer públics els objectius del PADICAT per al període 2006-2011²⁰, i se'n va publicar un primer balanç l'any 2008²¹, complementat amb un més recent, de l'any 2011²², que també en contempla els reptes de futur, destinats essencialment a consolidar i garantir la infraestructura del dipòsit, abordar la definició de les estratègies de preservació digital, crear l'hemeroteca digital a Internet, i cooperar amb altres arxius web i dipòsits de preservació digital per donar una resposta eficient als reptes de preservació digital i accés als recursos dipositats.

Pel que fa al balanç, s'establia en el moment inicial que l'objectiu genèric del PADICAT era dissenyar i produir un sistema que permetés que la BC compilés, processés i donés accés permanent a la producció digital catalana. Cinc anys després, l'objectiu genèric del projecte s'ha traduït efectivament en el disseny i la producció d'un sistema que ens permet actualment compilar, processar i donar accés a la part de la producció digital catalana a Internet que hem incorporat al dipòsit.

En un marc més operatiu, en la planificació del projecte s'assenyalaven tres eixos de treball que continuen vigents, atès que són característics dels models híbrids de captura:

En primer lloc, la compilació massiva dels recursos digitals publicats en obert a Internet. A partir d'una sèrie de captures de prova (2007-2008) del domini .cat, que van obligar a ampliar sensiblement els recursos destinats a captura i emmagatzematge, s'ha dut a terme una captura exhaustiva del domini corresponent a l'any 2009, dues més per a l'any 2010²³, i està programada la captura sistemàtica semestral.

Per tant, les pàgines web amb domini .cat es capturen i processen anualment. Complementàriament, es fa una captura semestral dels recursos procedents de les entitats que han signat convenis de col·laboració; una captura semestral dels recursos digitals procedents de recomanacions²⁴, i captures periòdiques de recursos que formen part dels monogràfics²⁵.

En segon lloc, l'impuls al dipòsit sistemàtic de la producció web de les entitats i les empreses de Catalunya. Des de l'inici del projecte, i amb l'objectiu de tancar 500 convenis de cooperació abans del final del 2011, s'han identificat fins a 2.000 institucions considerades agents principals de la producció digital catalana.

S'ha presentat el projecte a 1.800 d'aquests ens, i s'han formalitzat els 450 convenis de cooperació, amb una previsió per als propers mesos de complir l'objectiu de 500 entitats.

Finalment, la promoció de línies de recerca per mitjà de la presentació temàtica dels recursos digitals sobre determinats esdeveniments de la vida pública catalana. A partir de l'anàlisi de processos similars en altres projectes, i coincidint amb un calendari electoral regular, s'ha optat per efectuar una captura focalitzada d'un esdeveniment electoral anual relacionat amb campanyes electorals: al *Parlament de Catalunya* el 2006, les municipals del 2007, al *Congrés* i al *Senat* espanyol el 2008, al *Parlament Europeu* el 2009, i novament al *Parlament de Catalunya* el 2010.

Una acció de col·laboració amb l'*Escola Superior de Música de Catalunya* (ESMUC) va permetre ampliar aquesta oferta amb una nova fórmula: els recursos digitals catalans relacionats amb la música folk-rock.

Complementàriament, s'ha presentat un monogràfic dedicat als museus de Catalunya²⁶ i s'han fet captures ràpides d'esdeveniments a Internet, com el seguiment del debat de la prohibició dels toros al Parlament, l'editorial «La dignitat de Catalunya», «Zona9 música a la xarxa», les prèvies de la campanya *Jocs Olímpics Barcelona 2022* o els casals catalans a l'exterior.

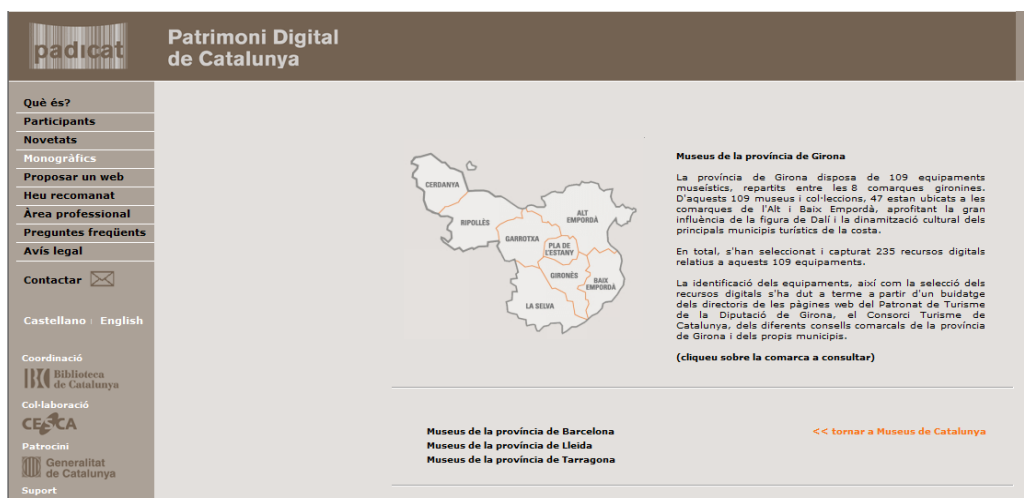


Fig. 2. PADICAT. Monogràfic de museus
(fuente: <http://www.padicat.cat/museus.php>)

A mode de conclusió, i el més important, sens dubte, és que s'està fent satisfactòriament un treball sistemàtic de compilació, processament i difusió del patrimoni digital de Catalunya a Internet.

L' instrument de preservació en alta seguretat de la *Biblioteca de Catalunya*

L'any 2009, la BC va crear un grup de treball intern de preservació digital, amb l'objectiu de dissenyar un repositori de preservació digital d'alta seguretat²⁷, que ha estat anomenat COFRE (*CO*nservem *per al Futur* *Recursos Electrònics*). Aquest grup va dedicar els primers esforços a tres tasques:

- Conèixer els fons digitals (o digitalitzats) de la BC objecte de preservació, quant a formats, nombre, ocupació de les dades i la seva previsió de creixement, metadades existents tant de digitalització com de preservació, suport d'emmagatzemament i estat actual de conservació i l'origen de creació de les dades (digitalitzacions internes, externes, documents nascuts digitals de productors externs, etc.).
- Establir els usos i funcionalitats del sistema. La BC va adoptar el model OAIS (*Open Archival Information System*)²⁸, que defineix les funcions, responsabilitat i organització d'un sistema de preservació. Així va poder determinar les parts que conformaran el sistema: una capa d'emmagatzemament que gestionarà els volums que contindran les dades, una capa de gestió que incorporarà les accions i rutines de preservació, la gestió i control d'accés i una capa d'ingesta que permetrà les càrregues pels diferents productors de continguts. Finalment es van identificar diverses necessitats relacionades amb els objectes digitals que calia afegir com a part de la capa de serveis: la difusió de documents a la Biblioteca o de manera oberta, depenent de si existeixen o no drets d'autor vigents. I la gestió de còpies d'alta qualitat per a ús comercial i d'investigació.

- Conèixer els programaris de preservació existents al mercat en aquells moment, tant de codi obert com de pagament, i determinar la seva capacitat d'adaptació a les necessitats de la BC. En aquest sentit no es va trobar cap programari que permetés una adaptació completa sense haver de sacrificar altres funcionalitats, raó per la qual finalment es va optar per fer un desenvolupament a mida per part del personal tècnic de la BC.

A partir de l'assoliment en les tres tasques, es va dissenyar un sistema que permetés abordar els elements següents:

- Gestió de continguts. Es va idear una estructura jeràrquica de contenidors que permet la seva nidació i el control d'accés a diferents nivells. Així mateix el model permet la creació de diverses versions del document permetent guardar còpies en diversos formats del mateix objecte, ja sigui per raons de l'ús de cada còpia o com versions cronològiques resultat de migracions de format o digitalitzacions posteriors. El sistema incorpora diversos sistemes de càrrega de continguts, tant de manera individual com càrregues massives de digitalitzacions de grans fons o de documents amb una estructura complexa de carpetes.
- Gestió de metadades descriptives, tècniques, administratives. El sistema permet mantenir diversos jocs de metadades i plantilles de càrrega amb valors de metadades predefinides. Es va optar per l'ús de *Dublin Core*²⁹ per a les metadades descriptives i administratives i adoptar un format propi "BC" per als números identificatius, les metadades tècniques i administratives pròpiament de preservació, basat en altres esquemes coneguts arreu del món com PREMIS³⁰ per a les dades administratives de preservació o MIX³¹, PBCore³² per a les metadades tècniques. La raó d'aquesta aposta, radica en el fet de no haver trobat un únic estàndard d'acceptació global si no diverses iniciatives recolzades per un o altre sector o comunitats, principalment en el marc de la descripció tècnica del format dels objectes: imatge fixa, àudio, audiovisuals, etc. No obstant és part del projecte la creació d'un mòdul d'extracció de metadades encapsulades en METS³³ i la creació d'un mapatge *BC-Premis*, BC-altres formats, per adaptar-se a les situacions que en el marc de metadades tècniques i de preservació es puguin trobar en un futur.
- Gestió de rutines de conservació i preservació. Un punt vital en qualsevol sistema de preservació és la integritat i autenticitat de les dades. En aquest sentit el sistema incorpora rutines cícliques de comprovació de la signatura digital dels fitxers i d'infecció per virus. Un altra punt important és l'accés a les dades, garantint-ne el seu accés ara i en un futur. El sistema realitza una detecció automàtica del format dels fitxers inclosos en la versió del document, equiparant-los amb estàndards internacionals (*Pronom UID*,³⁴ *MIME media type*³⁵), amb una doble finalitat: conèixer en tot moment les eines necessàries per a la lectura i recuperació dels fitxers; i conèixer els riscos i amenaces dels formats actuals que poden impossibilitar en el futur la seva lectura i recuperació

per a poder així planificar les estratègies necessàries per a impedir-ho: migració de les dades, conversió de format, etc.

Per elevar la fiabilitat del sistema respecte de la seguretat i autenticitat de les dades calia dotar el sistema amb una gestió d'usuaris als quals aplicar polítiques d'accés i rols que modelessin la seva capacitat d'actuació sobre els objectes de cada contenidor. Així mateix totes les accions dutes a terme sobre els objectes són enregistrades per a la seva consulta posterior.

Atès el volum total de dades existent i la previsió de creixement, els sistemes d'emmagatzemament que actualment té la BC no tenen capacitat suficient per suportar la càrrega total prevista. Per tant una gestió dels volums de dades flexible permetrà créixer segons les necessitats, incorporar altres volums de dades que s'adquireixin en un futur i facilitar una possible migració futura.

El sistema descrit correspon al que entenem com a repositori de preservació digital; ateses, però, les necessitats particulars de servei i gestió de la BC, s'ha previst el desenvolupament d'una sèrie de mòduls i funcionalitats addicionals que depassen l'estrictament considerat com a preservació digital.

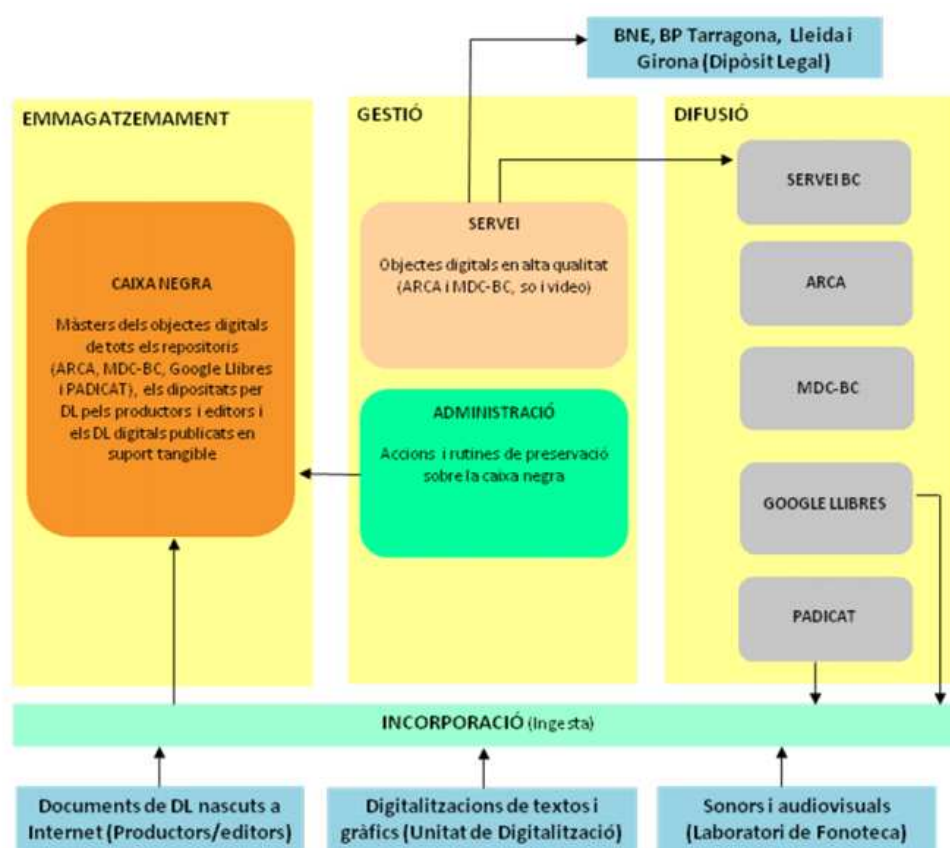


Fig. 3. BC. Esquema del repositori de preservació digital

Pel que fa a la gestió del DL de documents nascuts digitals, és objectiu del sistema comptar amb un mòdul específic, ubicat a la capa de servei, que permetrà als productors/editors, dipositar en línia els documents digitals en comptes de lliurar-ho en DVD o USB com fan actualment. La incorporació al repositori generarà les alarmes necessàries per a procedir a la comprovació des de les oficines del DL de la validesa dels objectes, la catalogació des del servei corresponent de la Biblioteca i la distribució a les biblioteques dipositàries corresponents: *Biblioteca Nacional d'Espanya*, Biblioteques Públiques de Girona, Tarragona i Lleida.

La BC, com hem explicat anteriorment participa i lidera als repositoris MDC i ARCA de difusió de documents que es troben en domini públic (lliures de drets) digitalitzats en baixa resolució, però també, en determinats casos, digitalitza documents amb drets d'autor vigents, que en conseqüència només poden ser distribuïts i consultats dins de la xarxa de comunicacions de la pròpia BC d'acord amb la legislació sobre propietat intel·lectual, i que en aplicació d'aquesta mateixa legislació presenten restriccions de còpia i reproducció. Per altra banda, regularment, la BC rep sol·licituds de reproduccions en alta qualitat a les que ha de donar resposta.

Tota aquesta casuística es resoldrà amb un mòdul de gestió de còpies d'alta qualitat i difusió dels continguts, enfocat a l'obtenció per part dels usuaris de còpies de documents subjectes a drets i difusió limitada a efectes de publicació o ús d'investigació. Aquest servei suposa la inclusió/integració de mecanismes d'acceptació legal de l'ús en fer les còpies, del compromís de pagament de drets i de pagament del servei.

El repositori de preservació estarà en producció dins del 2011 pel que fa a càrrega de continguts tant interna com per part dels productors/editors del Dipòsit Legal, la gestió de metadades i les rutines de preservació sobre els objectes digitals i es preveu desenvolupar els mòduls de consulta pública interna de documents amb drets vigents i la petició de còpies d'alta qualitat al llarg del 2012.

Dissenyar, implementar i mantenir sistemes de preservació digital és costós, tant des del punt de vista d'infraestructura tecnològica i programari, com d'expertesa i de personal especialitzat, tots dos aspectes no sempre a l'abast de qualsevol organització; és per tant un horitzó plausible que en el futur el repositori de preservació digital de la BC pugui esdevenir una eina de caràcter nacional i cooperatiu per a la preservació de col·leccions de l'àmbit català.

Conclusions

Hores d'ara ningú no fa qüestió de la necessitat de garantir l'accés al patrimoni digital, fent-ne difusió i assegurant la seva preservació. Les principals característiques i requeriments de la preservació digital comencen a ser conegudes en els àmbits professionals, tenim però poca pràctica real. Tampoc hi ha ara per ara una oferta comercial i diversificada de programaris de preservació digital que ens puguin ajudar a dur-la a terme; i encara que existís tindríem dificultats per avaluar-los a causa de la manca d'experiència.

Si el món digital és canviant, i això sí que ho sabem, la manera de fer preservació digital serà sens dubte un procés que caldrà adaptar i modificar de forma continuada. Els costos associats, encara que importants, probablement es reduiran amb el temps pel que fa a equipaments, però segur que s'incrementaran quant a les persones que s'hi dedicaran.

La Biblioteca de Catalunya ha planificat i executat tres eixos de treball que persegueixen garantir l'accés al patrimoni digital, així com la seva preservació. Dos dels vectors fan referència als continguts: la digitalització retrospectiva, i l'arxiu del patrimoni nascut digital i publicat a Internet. El tercer persegueix garantir la correcta preservació dels fons que conformen el patrimoni digital de la BC, nascut a partir dels processos de digitalització, o publicat a Internet per part de la societat catalana.

Però sabem que treballem en un escenari de futur incert, així que caldrà mantenir-se alerta dels canvis de formats, productes, distribució i tractament.

Amb el creixement exponencial d'informació digital, esdevindrà clau la selecció: no es podrà preservar tot, però sí que haurem de poder fer una bona fotografia de la producció digital, i esdevindrà també clau col·laborar, doncs la preservació digital es dibuixa d'una magnitud tal que ningú, sol, serà capaç d'afrontar.

És un futur estimulante, ple de reptes que ens han de motivar, i en el qual hem de superar les pors i resistències per mitjà de la confiança en la nostra pròpia experiència. Així és com ho entenem a la Biblioteca de Catalunya.

Referències bibliogràfiques

1. El febrer de 2011 el *Consell de Ministres* aprovava un projecte de *Llei sobre el Dipòsit Legal*. El text del projecte era el resultat del treball cooperatiu realitzat en el marc del *Consejo de Cooperación Bibliotecaria*, impulsat per la *Subdirección General de Bibliotecas del Ministerio de Cultura*, en el que hi són presents totes les comunitats autònomes. Vegeu-lo a:

ESpanya. CONSELL DE MINISTRES (1 abril 2011). "Proyecto de Ley 121/000118 de depósito legal". *Boletín Oficial Cortes Generales*, 118.1/2011. Disponible en:

<http://www.elderecho.com/actualidad/Proyecto-Ley-deposito-legal_EDEFIL20110405_0011.pdf>

[Consulta: 28 abril 2011].

2. Actualment l'arribada de documents electrònics a les oficines de DL depèn en bona mesura de la voluntat de l'editor o productor, al que a més, a dia d'avui, se li demana que faci un lliurament en suport tangible (paradoxal, si tenim en compte que són documents creats per ser utilitzats i distribuïts a la xarxa). La BC, de la que en forma part l'*Oficina del Dipòsit Legal de Catalunya*, ha contactat els darrers anys amb els principals editors i productors per sensibilitzar-los respecte a la necessitat de preservar aquests documents electrònics per a les generacions futures, i rep en DVD o USB, una part de la producció total. En l'actual redisseny dels processos que afecten a la preservació dels materials procedents de DL, es contempla l'establiment d'un mecanisme que permeti realitzar en línia tota la gestió dels documents electrònics arribats per dipòsit legal, de manera que els editors/productors puguin fer el lliurament dels documents via web i aquests s'incorporin, una vegada validats, al repositori de preservació digital que es presenta al final del present article. D'aquesta manera, tota la gestió de documents electrònics serà virtual, doncs la BC des del 2001 ja disposa d'una aplicació web que facilita la sol·licitud de números de DL i el lliurament de les dades dels documents per web. Quant als documents en suport físic, principalment CD's i DVD's d'enregistraments sonors i audiovisuals, la BC rep per dipòsit legal aproximadament 7.000 anuals, i en total es poden estimar en uns 50.000 els

que formen part de la seva col·lecció. Coneixedors de la degradació que pateixen aquests suports i de la seva esperança de vida, fins i tot en entorns de emmagatzemament òptims, es preveu incorporar aquests documents al repositori de preservació digital una vegada estigui en ple rendiment.

3. A partir del que consta en les lleis catalanes de biblioteques del 1981 (DOGC 123, 29/04/1981) i 1993 (DOGC 1727, 29/03/1993), la *Biblioteca de Catalunya* té com a missió recopilar, conservar i difondre la producció bibliogràfica catalana i la relacionada amb l'àmbit lingüístic català, a més, vetlla per la conservació i la difusió del patrimoni bibliogràfic.

Vegeu els documents d'estratègia:

BIBLIOTECA DE CATALUNYA (2004). *Pla estratègic de la Biblioteca de Catalunya 2004-2008*. Barcelona: BC. Disponible en:

<http://www.bnc.es/bc-/qualitat/pestrategic2004_2008.doc>

[Consulta: 28 abril 2011].

BIBLIOTECA DE CATALUNYA (2009). *Pla estratègic de la Biblioteca de Catalunya 2009-2012*. Barcelona: BC. Disponible en:

<http://www.bnc.es/bc/qualitat/pestrategic_2009_2012.pdf>

[Consulta: 28 abril 2011].

4. L'estratègia i els projectes, a excepció del *Google Llibres*, que va ser un acord posterior, foren presentats a:

LAMARCA, D.; SERRA, E. (2005). "L'estratègia de la Biblioteca de Catalunya en projectes digitals". En: *Ítem*, n. 41, pp. 41-43.

<<http://www.raco.cat/-/index.php/Item/article/view/40866/68116>>

[Consulta: 28 abril 2011].

5. Conegut mundialment com *Google Books*, a Catalunya en formen part la *Biblioteca de Catalunya*, que n'és la coordinadora, la *Biblioteca de l'Ateneu Barcelonès*, la *Biblioteca Pública Episcopal del Seminari de Barcelona*, la *Biblioteca del Centre Excursionista de Catalunya*, i la *Biblioteca de l'Abadia de Montserrat*. L'objectiu és possibilitar l'accés lliure a Internet de la còpia digital de prop de 200.000 documents lliures de drets.

<<http://books.google.cat>>

[Consulta: 28 abril 2011].

6. MDC (Memòria Digital de Catalunya) (2005). Biblioteca de Catalunya, Consorci de Biblioteques Universitàries de Catalunya. Disponible en:

<<http://mdc.cbuc.cat>>

[Consulta: 28 abril 2011].

7. ARCA (Arxiu de Revistes Catalanes Antiques) (2005). Biblioteca de Catalunya. Disponible en:

<<http://www.bnc.cat/digital/arca.index.html>>

[Consulta: 28 abril 2011].

8. Vegeu la memòria de l'anàlisi i el plantejament del projecte:

BIBLIOTECA DE CATALUNYA (desembre 2005). *Memòria del plantejament del projecte PADICAT (Patrimoni Digital de Catalunya)*. Barcelona: BC. Disponible en:

<<http://www.recercat.net/handle/2072/1757>>

[Consulta: 28 abril 2011].

9. Inclou els monogràfics de les campanyes electorals al *Parlament de Catalunya* el 2006, les municipals del 2007, al *Congrés* i el *Senat* espanyol el 2008, al *Parlament Europeu* el 2009, i novament al *Parlament de Catalunya* el 2010. Complementàriament, música folk-rock i museus de Catalunya. Vegeu-ho a:

<<http://www.padicat.cat/esdeveniments.php>>

[Consulta: 28 abril 2011].

10. Dades amb data 28 abril de 2011. Vegeu-ne les actualitzacions a:

<<http://www.padicat.cat/es/estadistiques.php>>

[Consulta: 28 abril 2011].

11. VIVES, J. (2005) "Aspectos de propiedad intelectual en la creación y gestión de repositorios institucionales". En: *El profesional de la información*, v. 14, n. 4, pp. 267-278. Disponible en:
<<http://www.elprofesionaldelainformacion.com/contenidos/2005/julio/4.pdf>>
[Consulta: 28 abril 2011].
12. Heritrix.
<<http://crawler.archive.org>>
[Consulta: 28 abril 2011].
Vegeu-ne un article explicatiu a:
MOHR, G. *et al.* (2004). "An introduction to Heritrix: an open source archival quality web crawler". En: *International Web Archiving Workshop*. Disponible en:
<<http://www.iwaw.net/04/Mohr.pdf>>
[Consulta: 28 abril 2011].
13. Arc File Format.
<[http://en.wikipedia.org/wiki/ARC_\(file_format\)](http://en.wikipedia.org/wiki/ARC_(file_format))>
[Consulta: 28 abril 2011].
14. NutchWax.
<<http://archive-access.sourceforge.net/projects/nutch>>
[Consulta: 28 abril 2011].
15. Hadoop.
<<http://hadoop.apache.org/core>>
[Consulta: 28 abril 2011].
16. Wayback.
<<http://archive-access.sourceforge.net/projects/wayback>>
[Consulta: 28 abril 2011].
17. Wera.
<<http://archive-access.sourceforge.net/projects/wera>>
[Consulta: 28 abril 2011].
18. Web Curator Tool.
<<http://webcurator.sourceforge.net>>
[Consulta: 28 abril 2011].
19. Vegeu-ne un article descriptiu a:
LLUECA, C. *et al.* (2010). "CAT (Curator Archiving Toll): millorant l'accés als arxius web" En: *International Internet Preservation Consortium meeting*. Disponible en:
<http://www.recercat.net/bitstream/2072/85525/2/Padicat_iipc_2010_CAT.pdf>
[Consulta: 28 abril 2011].
20. LLUECA, C. (2006). "El projecte PADICAT (Patrimoni Digital de Catalunya) de la Biblioteca de Catalunya". En: *10es Jornades Catalanes d'Informació i Documentació*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya. Disponible en:
<http://eprints.rclis.org/archive-/00006434/01/llueca_padicat.pdf>
[Consulta: 28 abril 2011].
21. CÓCERA, D.; LLUECA, C. (2008). "PADICAT: realitat i reptes de 3 anys d'arxiu web de Catalunya". En: *11es Jornades Catalanes d'Informació i Documentació*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya. Disponible en:
<http://eprints.rclis.org/archive/00013562-/01/llueca_padicat_jornades_2008.pdf>
[Consulta: 28 abril 2011].
22. LLUECA, C. *et al.* (2011). "El PADICAT, l'experiència catalana en l'arxiu d'Internet". En: *Lligall* [inèdit, previst tardor 2011].
23. El domini .cat tenia 31.125 registres actius durant la darrera captura massiva, el 2010, que es completen a efectes numèrics amb pàrquings, redreçaments i pàgines sense contingut, d'acord amb les dades públiques de la *Fundació puntCAT*.

24. El projecte promou la participació activa de l'usuari per mitjà de la recomanació de webs susceptibles de formar part de l'arxiu. Aquesta possibilitat, oberta a través d'un formulari, ha tingut un èxit considerable pel que fa a la participació dels usuaris (758 pàgines recomanades, en la data de presentació d'aquest article). No ha passat el mateix, però, en la rapidesa a l'hora de procedir a la captura d'aquests recursos, ja que s'han produït retards en el procés de captura i publicació.

25. L'any 2010 es van capturar de manera focalitzada 1.000 pàgines web per a la campanya electoral al Parlament a Internet, amb un volum aproximat de 4.000 captures.

26. Selecció i captures de 1.532 recursos digitals relatius als 657 museus i col·leccions museogràfiques de Catalunya. Vegeu el monogràfic a:

<<http://www.padicat.cat/museus.php>>

[Consulta: 28 abril 2011].

27. PÉREZ, K; SERRA, E. (2010). *Repositori de preservació digital de la Biblioteca de Catalunya: informe descriptiu i de situació*. Barcelona: Biblioteca de Catalunya. Disponible en:

<<http://www.recercat.net/handle/2072/97251?locale=ca>>

[Consulta: 28 abril 2011].

28. *Reference Model for an Open Archival Information System (OAIS)* (January, 2002). Consultative Committee for Space Data Systems, CCSDS 650.0-B-1, Blue Book, Issue 1. Disponible en:

<<http://public.ccsds.org/publications/archive/650x0b1.pdf>>

[Consulta: 28 abril 2011].

29. *Dublin Core Metadata Initiative*.

<<http://dublincore.org>>

[Consulta: 28 abril 2011].

30. *PREMIS (Preservation Metadata: Implementation Strategies)*.

<<http://www.loc.gov/standards/premis/>>

[Consulta: 28 abril 2011].

31. *MIX (NISO Technical metadata for Digital Still Images Standard)*.

<<http://www.loc.gov/standards/mix/>>

[Consulta: 28 abril 2011].

32. *PBCore, (Metadata standard for audiovisual media developed by the public broadcasting community)*.

<<http://pbcore.org/>>

[Consulta: 28 abril 2011].

33. *METS (Metadata Encoding & Transmission Standard)*.

<<http://www.loc.gov/standards/mets/>>

[Consulta: 28 abril 2011].

34. *PRONOM (on-line information system about data file formats)*.

<<http://www.nationalarchives.gov.uk/PRONOM>>

[Consulta: 28 abril 2011].

35. *MIME media types*.

<<http://www.iana.org/assignments/media-types/index.html>>

[Consulta: 28 abril 2011].