REPOSITORI DE PRESERVACIÓ DIGITAL DE LA BIBLIOTECA DE CATALUNYA

Informe descriptiu i de situació

Redactat per Karibel Pérez i Eugènia Serra
del Grup de Preservació Digital de la BC

Barcelona, desembre 2010

RESUM

Com a responsable de la preservació del patrimoni bibliogràfic a Catalunya, la Biblioteca de Catalunya ha dedicat molts recursos a digitalitzar documents analògics, alhora que ha incorporat al seu fons documents creats directament en format digital.

En l'actualitat representen un volum important, i per això el Grup de Preservació Digital de la BC ha descrit les necessitats i processos que són la base del disseny d'un repositori que garanteixi la perdurabilitat d'aquests documents. El present informe en descriu els detalls.

1 INTRODUCCIÓ

La Biblioteca de Catalunya en tant que biblioteca nacional té la responsabilitat de preservar el patrimoni bibliogràfic.

Amb el canvi de mil·lenni, les biblioteques nacionals varen iniciar un procés d'evolució del seu model per donar servei a tots els seus clients potencials, presencials o no. La BC va aprovar el 2004 i va revisar el 2009 un pla estratègic[1] amb la visió de ser una biblioteca oberta, fiable i orientada a l'usuari, endegant iniciatives orientades a la preservació del patrimoni tant en l'entorn analògic com digital[2]. En l'entorn analògic, a més de les accions tradicionals que ja duia a terme, ha fet un esforç important de digitalització per reduir l'ús dels originals analògics, contribuint així a la seva perdurabilitat.

La digitalització ha general un volum important i creixent de substituts digitals d'originals analògics que cal preservar perquè continuïn actuant com a còpies d'accés, i a la vegada perquè constitueixen en sí mateixos fons patrimonials que cal conservar per les generacions futures.

La biblioteca vol liderar i impulsar accions de preservació cooperatives a nivell de país per donar resposta a una preocupació que arxius, biblioteques, museus i institucions patrimonials en general han fet palesa en els darrers anys. Trobar una solució o si més no iniciar un plantejament global i alineat amb les tendències internacionals és estratègic ara i ho serà en els propers anys.

Amb aquesta visió àmplia, la Biblioteca es troba amb diverses vies d'adquisició de documents digitals: proveïdors editorials que volen dipositar documents creats a Internet, d'acord amb la llei vigent de dipòsit legal, webs procedents del seu propi repositori PADICAT d'arxiu de la web catalana, objectes digitals creats per la mateixa BC mitjançant la digitalització i documents nascuts digitals

[1] A partir del que consta a les lleis catalanes de biblioteques de 1981 (DOGC 123, 29/04/1981) i 1993 (DOGC 1727, 29/03/1993), la Biblioteca de Catalunya té per missió recopilar, conservar i difondre la producció bibliogràfica catalana i la relacionada amb l'àmbit lingüístic català, i vetlla per la conservació i la difusió del patrimoni bibliogràfic. Vegeu els documents d'estratègia: Biblioteca de Catalunya. *Pla estratègic de la Biblioteca de Catalunya 2004-2008*. Barcelona: BC, 2004. <http://www.bnc.es/bc/qualitat/pestrategic2004_2008.doc> [Consulta: 21/12/210]; i Biblioteca de Catalunya. *Pla estratègic de la Biblioteca de Catalunya 2009-2012*. Barcelona: BC, 2009. <http://www.bnc.es/bc/qualitat/pestrategic_2009_2012.pdf> [Consulta: 21/12/210].

[2] Lamarca, Dolors, Serra, Eugènia. "Deu pinzellades de la història recent de la Biblioteca de Catalunya (1993-2007)". En: *Item : revista de biblioteconomia i documentació*, núm. 46 (2007), p. 35-51. <http://www.raco.cat/index.php/Item/article/view/40866/68116> [Consulta: 21/12/210].

Lamarca, Dolors, Serra, Eugènia. "L'estratègia de la Biblioteca de Catalunya en projectes digitals". En: *Item : revista de biblioteconomia i documentació*, núm. 41 (2005), p. 41-53. <http://www.raco.cat/index.php/Item/article/viewFile/123057/170807> [Consulta: 21/12/210].

i publicats en CD o DVD, principalment sonors i audiovisuals que s'han incorporat al llarg dels anys a la BC majoritàriament per dipòsit legal.

La preservació d'aquest patrimoni requereix doncs del disseny i creació a la BC d'un dipòsit d'alta seguretat, un repositori de preservació.

En el cas de la BC cal tenir en compte que partim d'una situació existent amb repositoris digitals de difusió en producció, fet que condiciona el desenvolupament actual del repositori de preservació en el sentit que no cal preveure un mòdul de difusió dels documents per Internet; per altra banda, som conscients que a més dels continguts descrits en aquest document existeixen d'altres recursos a preservar ubicats a repositoris cooperatius, institucionals i d'empreses, als quals en una segona fase caldrà proporcionar solucions, a partir de l'establiment de protocols i procediments que els regulin com a repositoris de preservació segurs.

Finalment, afegir que en un futur, una vegada s'hagi consolidat el repositori de preservació i s'hagi avaluat el seu funcionament, no es descarta la possibilitat que pugui esdevenir un repositori cooperatiu.

2 METODOLOGIA

La Biblioteca de Catalunya conscient de la necessitat d'avançar-se a situacions de pèrdua d'informació digital va crear l'any 2008 el Grup de Preservació Digital de la BC, de caràcter transversal, amb l'objectiu d'analitzar l'estat de la qüestió, identificar les necessitats i dibuixar les accions a realitzar per a garantir la perdurabilitat dels objectes i continguts digitals dels seus dipòsits.

El Grup està format per Karibel Pérez i Ramon Novoa (Àrea de Tecnologia de la Informació), Paquita Navarro (Unitat de Digitalització), Margarida Ullate (Fonoteca), Sergi Font (projecte Google), Ciro Llueca (Projecte PADICAT) i Eugènia Serra (Coordinació General).

La composició del grup respon a la conveniència de treballar conjuntament els diferents perfils i persones implicades, des dels creadors d'objectes als que hauran d'utilitzar el sistema per dipositar-ne còpies.

El grup va recollir la informació dels continguts a preservar, les metadades utilitzades, els formats existents i el volum dels continguts. En definitiva, es va construir el catàleg de fitxers. A partir d'aquesta informació s'ha definit l'organització dels continguts en funció del procediment d'entrada d'aquests al sistema i del seu ús, les metadades a aplicar, i els tipus d'usuaris i les polítiques de drets; paral·lelament s'han analitzat les opcions de programaris per gestionar el sistema, amb el resultat final d'optar per un desenvolupament propi; aquest punt es troba explicat en detall més endavant.

3 USOS

La Biblioteca de Catalunya, com a conseqüència de la seva activitat de preservació i difusió del patrimoni català, té identificades com a necessitats de servei en relació amb els objectes digitals, tres tipologies:

3.1 Difusió

3.1.1 *Documents en domini públic o amb dret de comunicació pública*.

Es tracta de reproduccions digitals en baixa resolució d'originals analògics que es difonen actualment mitjançant els repositoris digitals ARCA[3], MDC[4] i *Google Llibres*[5].

3.1.2 *Documents amb drets d'autor vigents*.

Són documents que només es poden consultar presencialment a la Biblioteca ja que no es permet la comunicació pública a Internet (és el cas dels enregistraments sonors analògics que s'han digitalitzat per motius de preservació). També entrarien en aquest apartat els documents nascuts digitals i publicats en suports com CD's i DVD's arribats a la BC principalment per dipòsit legal.

3.2 Servei i ús comercial

Obtenció per part dels usuaris de còpies tant de documents en domini públic com 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.

3.3 Preservació

Aplicació de les rutines i accions que garanteixin la pervivència dels objectes digitals dipositats al sistema.

[3] *ARCA (Arxiu de Revistes Catalanes Antigues)* <http://www.bnc.cat/digital/arca/index.html>
[4] *MDC (Memòria Digital de Catalunya)* <http://mdc.cbuc.cat/>
[5] Els documents de la BC digitalitzats i accessibles actualment es poden consultar a *Llibres de la BC a Google Llibres* <http://www.bnc.cat/digital/google/cerca.php>

4 DESCRIPCIÓ DEL CONTINGUT A PRESERVAR

Els documents digitals que es preservaran a la BC són fruit de la seva activitat en diversos àmbits. La naturalesa de cada activitat, els mecanismes de creació i captura, així com les vies d'arribada a la BC dels documents donen com a resultat objectes digitals amb característiques diferenciades.

4.1 Contingut

4.1.1 Documents digitalitzats per la BC d'originals analògics

Des de l'any 2005 la BC dedica un pressupost anual més o menys ampli a la digitalització dels seus fons d'acord amb els següents criteris:

- que siguin de domini públic (sense drets o amb autorització dels titulars)
- que presentin un grau de fragilitat elevat que no recomani la consulta per risc de fer-se malbé. Per exemple els discos de pedra només es poden consultar una vegada digitalitzats.
- que presentin requeriments de consulta especials. Per exemple les plaques de vidre requereixen d'una taula de llum.
- que tinguin un estat físic adequat (viables de ser digitalitzats perquè les avantatges de disposar de la còpia digital per a consulta supera el risc de la manipulació i l'exposició a la llum en digitalitzar el document).
- que coincideixin amb els objectius dels dipòsits digitals que lidera o en els quals col·labora la biblioteca.
- que siguin d'interès per als investigadors. Les col·leccions singulars i diferenciadores, especialment els documents manuscrits i impresos, textuals i gràfics publicats abans de 1800 en el cas de monografies i anteriors a 1936 en el cas de publicacions periòdiques.
- que formin part dels "tresors" de la col·lecció.

Fruit d'aquesta tasca es disposa a 2010 de més de 40.000 documents digitalitzats que representen més de 4 milions d'imatges d'alta i baixa resolució més el text de l'OCR (en termes d'ocupació d'espai suposa 3 TB d'imatges de difusió en JPEG + 65,25 TB d'imatges de preservació en TIFF); i prop de 2.000 hores de so i vídeo.

En tots els casos es disposa d'un màster i una còpia d'accés emmagatzemat en servidors, discs durs externs o DVD; addicionalment, dels que són de domini públic o es disposa de dret de comunicació pública, hi ha una segona còpia de difusió emmagatzemada al dipòsit digital d'accés ARCA (Arxiu de Revistes Catalanes Antigues) o MDC (Memòria Digital de Catalunya).

4.1.2 Documents digitalitzats per la BC dins del projecte Google Llibres

Com a soci del projecte Google Llibres, la BC ha digitalitzat més de 45.000 monografies que a 2010 es troben disponibles a Internet, amb previsió de créixer fins a superar els 60.000 documents digitalitzats. En aquest cas la còpia d'accés és la que publica Google a la seva plataforma i la BC guarda una còpia a efectes de preservació. El format d'aquests fitxers és JPEG2000.

4.1.3 Documents nascuts digitals a Internet

El 2005 la BC posava en marxa el repositori PADICAT (Patrimoni Digital de Catalunya)[6]. PADICAT es basa en l' aplicació d'una sèrie de programes informàtics que permeten la captura, l'emmagatzematge, l'organització, la preservació i l'accés permanent a les pàgines web publicades a Internet considerades patrimoni català.

A finals de 2010 PADICAT compta amb 39.401 pàgines webs dipositades, que comporten 107.804 captures (versions de webs) i 240 milions de fitxers, amb una ocupació de 7,5 TB d'espai. Els formats dels fitxers són en aquest cas heterogenis si bé, predominen quatre formats sobre la resta, 84% d'html/txt, 10% entre gif i jpeg i entorn a l'1% de pdf[7].

Addicionalment, via dipòsit legal, els productors comencen a fer dipòsit de llibres i revistes electròniques creades en línia, en suport DVD. De moment és un volum molt petit però és preveu un creixement important per als propers anys.

4.1.4 Documents nascuts digitals publicats en suports tangibles

La BC rep per dipòsit legal aproximadament 100.000 documents a l'any, d'aquests uns 7.000 anuals corresponen a documents digitals, principalment sonors i audiovisuals. Els suports d'aquests documents, CD o DVD principalment, té una esperança de vida a l'entorn dels 25 anys[8], depenent de les condicions d'emmagatzemament. A diferència dels documents tradicionals en paper, el fet que un CD o DVD es faci malbé implica en molts casos la impossibilitat d'accedir a la informació que contenen i, per tant, la pèrdua total d'aquesta.

4.2 Quadre resum

Recull la relació de tipus de documents digitals creats o capturats per la BC fins a 2010; no inclou per tant els documents descrits a l'apartat 3.1.4 ni els

[6] *PADICAT (Patrimoni Digital de Catalunya)* <http://www.padicat.cat/>
[7] Radiografies actualitzades a *Què tenim?:* <http://www.padicat.cat/estadistiques.php>
[8] *Preservation Management of Digital Materials: The Handbook*. London: Digital Preservation Coalition, 2008 <http://www.dpconline.org/advice/preservationhandbook>

documents dipositats per productor/editors de DL dels documents digitals nascuts a Internet.

La informació que es descriu per a cada tipus és el format de les còpies d'accés i de preservació, les metadades tècniques i de preservació que a priori –abans de començar a desenvolupar el repositori de preservació- la BC va considerar que calia recollir en previsió d'accions de preservació futures, i els suports en que s'han estat guardant els fitxers fins ara.

Tipus de document/ suport	Format preservació	Format accés	Metadades tècniques i de preservació	Suport d'emmagatzemament previ al repositori
Revistes i diaris per ARCA	TIFF	JPEG TXT	Tipus de recurs i format	DVD i Disc durs externs
Llibres anteriors a 1801 per a MDC	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	DVD/CD i Disc durs externs
Llibres posteriors a 1800 per a MDC	TIFF	JPEG TXT	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	DVD/CD i Disc durs externs
Llibres per a Google Llibres	JPEG2000	JPEG2000	Format, productor, característiques físiques, història de canvis	Servidor BC
Manuscrit	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	DVD/CD i Disc durs externs
Incunables	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	DVD/CD i Disc durs externs
Partitures	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	DVD i Disc durs externs
Mapes	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	Discs durs externs i DVD
Cartells	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	Discs durs externs i DVD

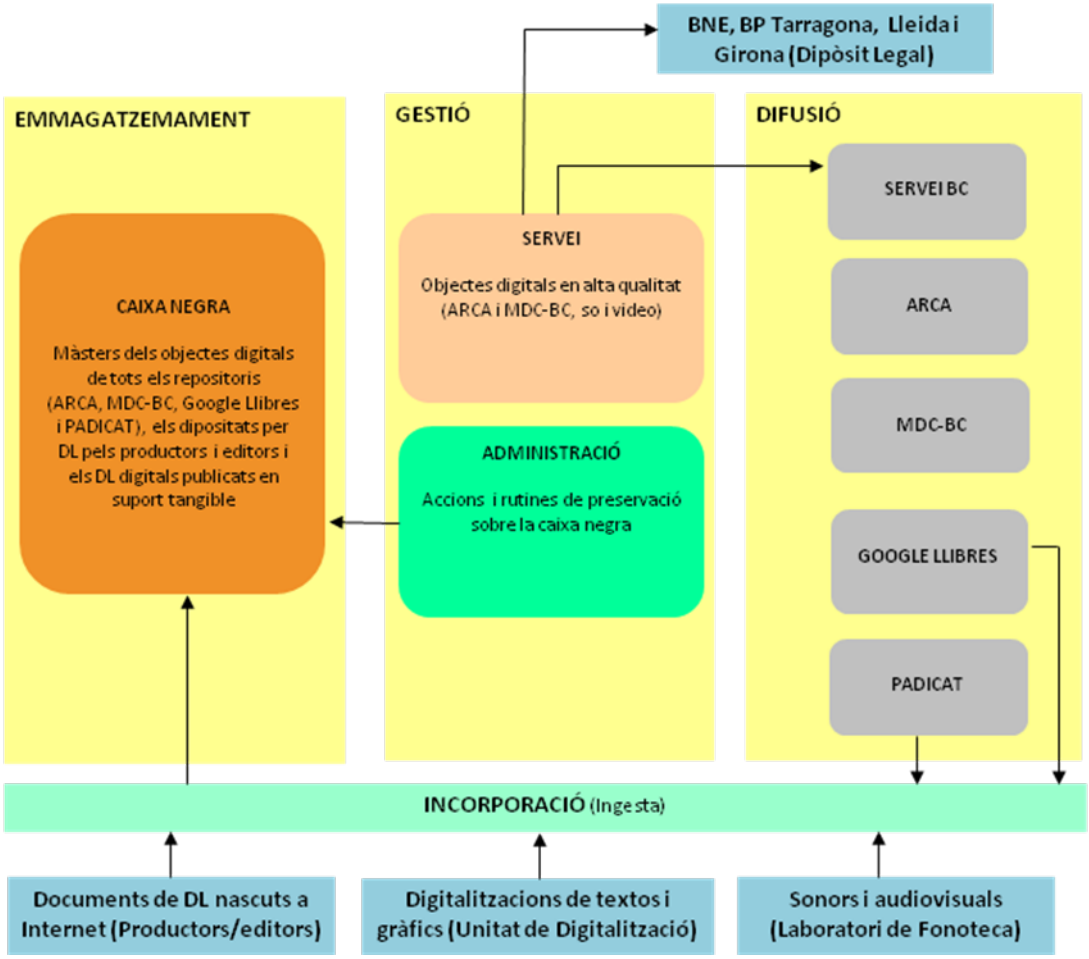
Suports Fotografia	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis canvis	Discs durs externs i DVD
Materials gràfics	TIFF	JPEG	Tipus de recurs, format, productor, resolució, compressió, definició, característiques físiques i història de canvis	Discs durs externs i DVD
Sonors	WAV o BWF	MP3	Tipus de recurs, format, profunditat de bits, freqüència de mostreig, aparells usats per a reproducció (model), corba d'equalització, velocitat, convertidor, nombre de canals, empresa i operador.	Servidor BC
Audiovisuals	MPEG-2 Vídeo MPEG-1 Àudio	MPEG-2 Vídeo MPEG-1 Àudio	Tipus de recurs, format, compressió de so i vídeo, targeta gràfica, aparells de reproducció, data de captura i empresa.	Servidor BC
Recursos web	TXT, HTML, JPEG, GIF PDF		Identificador a l'arxiu, URL del recurs capturat, data inici captura, data darrera modificació, format, estatus del fitxer Content checksum, HTTP Capçalera que identifica el "robot" que l'ha capturat.	Servidor CESCA

5 MODEL

El model adoptat per a la preservació és el model OAIS (Open Archival Information System)[9]. OAIS és un model conceptual de gestió, emmagatzemament i preservació a llarg termini de dades digitals. El model fou elaborat per la Consultative Committee for Space Data Systems, i el 2003 va ser acceptat com a norma ISO (ISO 14721:2003)[10]. Defineix a grans trets les funcions, responsabilitats i organització d'un sistema de preservació.

A aquest model cal afegir les capes específiques que la BC requereix de servei i ús comercial, i dipòsit legal.

L'esquema complet del sistema atenent als usos descrits prèviament queda recollit en l'esquema següent



[9] *Reference Model for an Open Archival Information System (OAIS),* Consultative Committee for Space Data Systems, CCSDS 650.0-B-1, Blue Book, Issue 1, January 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>

[10] ISO 14721:2003 *Space data and information transfer systems -- Open archival information system -- Reference model* <http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683>

6 MAQUINARI

6.1 Maquinari i volums de dades

A finals de 2009 la BC va adquirir una infraestructura que permetrà realitzar primer una prova pilot i després una vegada avaluada la prova, posar el sistema en producció i començar a traspassar-hi els fitxers que actualment es troben emmagatzemats a servidors locals, discos durs, DVD, i els que incorporin proveïdors/editors per DL. Es tracta d'un Clúster NetApp FAS3140 de doble Controladora (activa-activa) que permet créixer fins a 420TB en brut. En configuració RAID-DP + Hot Spare s'assegura la disponibilitat del sistema i la recuperació de les dades davant de la fallida de fins a 2 discos al mateix temps, sense haver de recuperar d'altres còpies. La capacitat actual neta és de: 30TB.

La BC a més disposava des de l'any 2006 d'un HP StorageWorks 4000 Enterprise Virtual Array de doble controladora com a sistema d'emmagatzemament central de la BC. La capacitat actual neta disponible per a volums de dades per al Repositori és de 18 TB.

Sumant tots dos magatzems de dades disposem actualment de 48TB nets.
Al llarg de 2011-2012 s'anirà ampliant la capacitat per cobrir la totalitat de necessitats d'emmagatzemament de la BC.

L'aplicació resideix en un servidor Linux de doble processador amb 6 GB de memòria RAM.

6.2 Balanceig dels volums de dades

Donada la gran quantitat de fitxers a emmagatzemar, aquests es desaran en diversos volums de dades, que podran anar creixent o afegir-ne de nous segons les necessitats.

En principi el tipus d'emmagatzemament serà via NFS.
S'ha creat un algoritme que permet per programa, i en el moment de la càrrega, seleccionar en quin volum de dades es guardes els fitxers corresponents a cada versió del document amb l'objectiu de balancejar automàticament la càrrega dels volums. Aquest algoritme es basa en els següents criteris:

- No es separaran els fitxers d'una mateixa versió del document.
- Poden existir versions diferents d'un mateix document en diversos volums de dades.
- Ha de permetre canviar les dades de volum d'una forma ràpida i senzilla, sense que aquest canvi suposi una nova càrrega de dades en l'aplicació.

Algoritme usat:

- Es descarten els volums de dades amb espai disponible inferior a l'espai necessari per a guardar els fitxers de la versió d'un document.
- Es calcula el ràtio entre l'espai disponible en el volum respecte del total.

- Es calcula el ràtio entre l'espai necessari per a desar els fitxers de la versió d'un document respecte a l'espai disponible en el volum.
- Es ponderen aquests ràtios i el volum que millor puntuació treu, guarda els fitxers.

7 PROGRAMARI

L'any 2010 s'han avaluat programaris per gestionar repositoris per determinar la viabilitat d'utilitzar-los, atenent a les necessitats de la BC.

D'una primera valoració es va considerar que Dspace podia ser una solució viable, i es va procedir a instal·lar la versió 1.6, adaptar el procediment de càrrega manual, crear els scripts de càrregues massives i inserir els jocs de metadades adoptats per la BC. Finalment es va concloure que, la versió disponible en aquells moments, no s'adaptava a les necessitats de la BC en matèria de preservació digital: mancaven rutines de preservació (com la identificació del format dels fitxers en el moment de la càrrega), presentava limitacions en els jocs de metadades (com la qualificació de segon nivell, necessària en esquemes com PREMIS[11]), i que caldria desenvolupar els mòduls específics de servei i dipòsit de documents per DL. Es descarta el seu ús.

Entre els programes comercials que comencen a sorgir es considera Rossetta comercialitzat per l'empresa Exlibris. Rossetta es perfila com una solució completa per a la preservació digital, però es tracta d'un producte de pagament i l'evolució de disponibilitat econòmica de la BC no permet abordar una despesa d'aquestes característiques. Afegir que, els mòduls específics de servei i dipòsit de documents per DL igualment s'haurien de desenvolupar. Es descarta la seva adquisició.

Finalment s'opta per realitzar un desenvolupament a mida per part de l'Àrea de Tecnologia de la Informació de la BC; es fa sobre una plataforma de servidor web Apache 2 i base de dades PostgreSQL 8. L'aplicació es desenvolupa amb PHP 5.3.

[11] *PREMIS (Preservation Metadata: Implementation Strategies).* <http://www.loc.gov/standards/premis/>

8 ELEMENTS DEL SISTEMA

El sistema base inclourà:

- Estructura de carpetes/contenidors, definida prèviament, per a guardar físicament els documents, físicament.
- Base de dades que inclou:
  - Gestió metadades descriptives, tècniques, administratives.
  - Gestió rutines conservació i preservació.
  - Gestió usuaris i seguretat.
  - Gestió de continguts: càrregues.
  - Gestió del Dipòsit Legal de documents nascuts digitals.
  - Gestió de còpies.
- Caixa negra.

8.1 Estructura de contenidors

Els continguts s'organitzen en contenidors en funció de la via de dipòsit:

- Digitalitzacions
- Dipòsit legal
- Recursos web
- Documents nascuts digitals en suport físic.

Cada contenidor s'estructura si cal en un segon nivell de contenidors, és el cas del contenidor Digitalitzacions que conté:

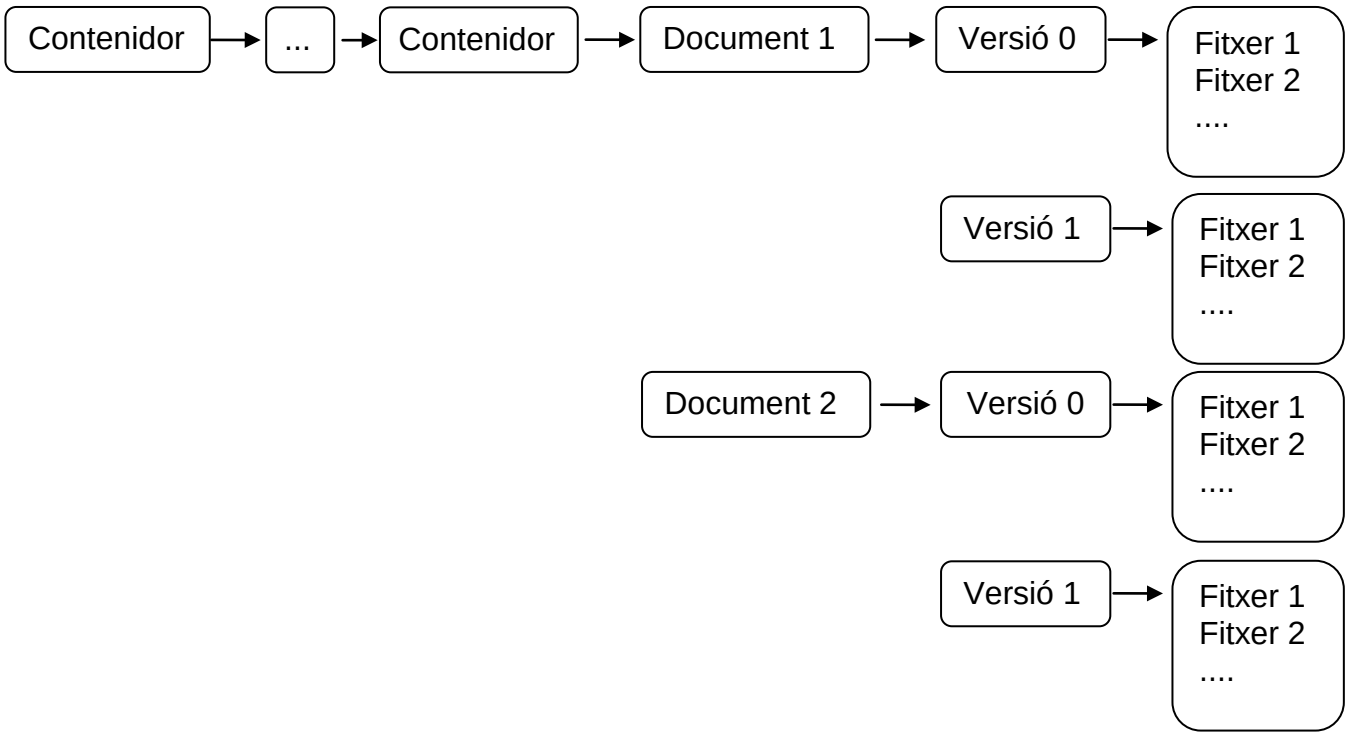
- Monografies
- Google Llibres
- Gràfics
- Publicacions periòdiques
- Sonors
- Audiovisuals

Depenent de la mateixa estructura del document original es crearan més nivells de contenidors, això és especialment útil en determinats casos, com per exemple, les publicacions periòdiques, on cada número de la publicació constitueix un document amb les seves versions i esdeveniments, que s'agrupen a un contenidor que correspon al títol de la publicació.

Aquesta organització facilita l'establiment de perfils d'usuaris diferents amb més o menys drets sobre cada nivell de l'estructura.

Els documents s'organitzen doncs dins del segon o següents nivells de contenidors. Cada document conté una o diverses versions, per exemple, una versió seria la còpia en TIFF i una altra podria ser la còpia en PDF.

Cada versió inclou els diferents esdeveniments que s'hi associen, d'acord amb les recomanacions de PREMIS.



8.2 Base de dades

8.2.1 Gestió de metadades

La Biblioteca de Catalunya utilitza als seus repositoris l'esquema Dublin Core, i també pel repositori de preservació pel que fa a les metadades descriptives i administratives. Quant a les tècniques i de preservació s'ha definit un joc de metadades propi, si bé el sistema disposarà d'un mapatge BC-PREMIS. Així mateix s'ha decidit finalment fer servir el mapatge BC per a crear les metadades que incorporaran identificadors del document.

Metadades	Estàndard	A nivell de...
Descriptives bibliogràfiques	Dublin Core / BC	Document
Administratives (drets)	Dublin Core	Document
Tècniques	BC	Versió i Fitxers
Preservació	BC	Versió

8.2.1.1 Metadades descriptives i administratives

Els elements descriptius s'han reduït al mínim, donat que s'inclou l'identificador del registre bibliogràfic del catàleg en línia (que hi enllaça) on s'hi troba tota la descripció. D'aquesta manera s'evita la redundància de feina i informació.

En crear el document el sistema assigna la data de creació i demana al depositant que seleccioni el contenidor, assigni l'estat del procés en que es troba i es marqui si el document és de domini públic.

Metadada	Especificacions	Obligatòria
Descriptives i administratives		
bc.identificador.bibliografic	*Núm. de registre bibliogràfic al catàleg de la BC (actualment Millennium)*	**NO**
bc.identificador.exemplar	*Núm. de registre exemplar al catàleg de la BC (actualment Millennium)*	**NO**
bc.identificador.topografic	*Topogràfic de l'exemplar*	**NO**
bc.identificador.codibarres	*Codi de barres de l'exemplar*	**NO**
bc.identificador.numDL	*Núm. de Dipòsit Legal*	**SÍ pel contenidor DL**
dc.creator	Entitat (una persona, una organització o un servei) que té la responsabilitat principal de la creació del contingut del recurs En la mateixa forma del catàleg	**NO**
dc.title	*Títol donat al document* Doneu el títol principal i, si és el cas el subtítol, en aquest element. Independentment de la tipografia usada en el document, doneu el títol en minúscula (excepte inicials) i amb accents (si n'hi ha).	**SÍ**
dc.source	*Referència a un recurs del qual se'n deriva el recurs present* L'element Source s'utilitzarà per citar el document del qual deriva el recurs digital. Es donarà en la següent forma: - Documents Impresos: Publicació original: Peu d'impremta - Publicacions Periòdiques: Publicació original: Peu d'impremta, Numeració - Documents manuscrits: Document original: Biblioteca de Catalunya. Topogràfic	**NO**
dc.description	*Descripció del contingut*	**NO**
dc.rights	*Estats dels dret.* Valors possibles: - Domini públic - Drets vigents	**SÍ**
dc.accesRights	Condicions d'ús i reproducció del document *Ex.: Accés i còpia permesa amb finalitat d'estudi o recerca, citant la font "Biblioteca de Catalunya". Per a qualsevol altre ús cal demanar autorització*	**NO**
dc.rightsHolder	*Propietari del copyright*	**NO**

8.2.1.2 Metadades tècniques i de preservació

Metadada	Especificacions	Obligatòria	Valors comuns
Tècniques			
bc.audio.format	*Format compressió d'àudio*	**SÍ**	wav,mp3,aiff

bc.audio.frequenciaMostratge	*Freqüència de mostratge,* ex.: 96kHz	**SÍ**	44.1 kHz,48 kHz,96 kHz 192 kHz
bc.audio.profunditatdeBits	*Profunditat de bits en àudio indicadora del rang dinàmic, ex.*: 24 bits	**SÍ**	24 bits, 1 bits, 8 bit
bc.audio.durada	*Durada del document,* hh:mm:ss	**NO**	
bc.audio.taxadeBits	*Taxa de bits d'àudio,* ex.: mono 192 kbps, estèreo 384 kbps	**NO**	
bc.audio.canals	*Nombre de canals d'àudio,* ex.: 2 canals (estèreo)	**NO**	1 (monofònic) 2 (monoaural) 2 (estèreo)
bc.aplicacio.format	*Format aplicació* Per a documents generats amb paquets d'ofimàtica,disseny gràfic, etc.	**SÍ**	pdf, msword, acad
bc.aplicacio.versio	*Versió*	**NO**	
bc.aplicacio.aplicacio	*Aplicació amb la qual s'ha generat el document*	**NO**	
bc.imatge.format	*Format i compressió d'imatge* *Indicació de la compressió i, si n'hi ha, del seu nivell,*	**SÍ**	TIFF sense compressió JPEG, compressió alta JPEG, compressió baixa
bc.imatge.resolucio	*Nombre de píxels usats per representar la imatge, expressats en dimensions de píxel, ppp...,*	**SÍ**	300 ppp, 400 ppp, 600 ppp
bc.imatge.profunditatdeBits	*Definició del color en imatges,*	**SÍ**	1 bit, 8 bit, 24 bits
bc.imatge.canals	*Nombre de canals d'imatge*, ex.: 3 en mode RGB, 1 en b/n o escala de grisos, 4 en CMYK	**NO**	*3,1,4*
bc.text.format	*Format del text*	**SI**	txt, html, xml
bc.text.versio	*Versió del format*	**NO**	
bc.text.codificacio	*Codificació del text*	**NO**	utf-8, latin-1
bc.video.format	*Format compressió de vídeo*	**SI**	mpeg2,mpeg4
bc.video.taxadeBits	*Taxa de bits de vídeo,* ex.: 1930 kbps	**NO**	
bc.video.fotogramaMida	*Mida del fotograma,* ex.: 720x576px	**SÍ**	
bc.video.fotogramesperSegon	*Fotogrames per segon,* ex.: 25.00fps	**SÍ**	
bc.video.relacioAspecte	*Relació d'aspecte,* ex.: (4:3)	**NO**	
bc.video.durada	*Durada del document,* hh:mm:ss	**NO**	
De preservació			
bc.event.descripcio	Descripció de l'esdeveniment	SI	Digitalització, Migració
bc.event.data	*Data de l'esdeveniment, ex.: 20080622, 200911, 20100115-20100311, 2007* ** Data única o interval de dates si el conjunt del document s'ha digitalitzat en dates diferents*	**NO**	
bc.event.productor	*Agència responsable de la creació física del recurs, ex. Biblioteca de Catalunya, Fonotrón* ** Es pot fer constar el nom/s de l'enginyer/s a continuació del nom de l'agència.*	**NO**	Biblioteca de Catalunya Archygest Artyplan Docout ADHOC Proco S.A. Docuteca ...

bc.event.dispositiu	*Dispositiu de captura o programari de migració,...* *Indicació de la marca i model del dispositiu de captura, ex.: Metis DRS A1, Archeophone, 160 rpm*	**NO**	Nikon D2-X Zeutschel OS 14000 A0 Metis DRS A1 Zeutschel OS 10000TT Bookeye BE3-SCL-R1 Epson Expression 10000 XL
bc.event.notes	*Notes addicionals de l'esdeveniment*	**NO**	

8.2.2 Gestió de rutines i accions de preservació

Per garantir la pervivència dels documents/objectes tal com foren creats cal assegurar el manteniment d'una sèrie de qualitats: fixesa, integritat i autenticitat. Aquestes qualitats s'aconsegueixen mitjançant l'aplicació d'una sèrie de rutines i praxis. El repositori de preservació digital de la BC aplica les següents:

Accions i rutines de preservació	Eines utilitzades	Acció	Indicadors
Identificació del format del fitxer	DROID Libmagic	En el procés de càrrega	Pronom UID (PUID) Descripció del format i mimetype
Comprovació de virus	ClamAV	En el procés de càrrega i posteriorment	Infecció/virus
Integritat dels fitxers *(verificar que el fitxer no ha estat modificat des que es va dipositar al repositori i que és llegible).*	Llibreria php	Comprovació cíclica sobre tots els fitxers	Checksum MD5 original i calculat periòdicament
Autenticitat (*equip reduït i control•lat de persones que puguin accedir a les còpies de seguretat, aplicant rutines que permetin veure el rastre de quan, qui i quines accions es fan a cada fitxer de preservació)*	Sistema control·lat d'usuaris i drets	Guardar i analitzar logs diàriament.	Informes periòdics de modificació i supressió de fitxers i documents
Seguretat de la xarxa	Xarxa interna Firewalls	Els volums de dades no són accessibles fora de la xarxa de la BC. El servei a l'usuari extern es donarà via serveis web. Control per IP + LDAP per a l'accés intern a l'aplicació de gestió del Repositori.	
Còpies de seguretat (backups) *(per a assegurar la recuperació de les dades en cas de desastre greu en les instal·lacions de la BC que provoquin la pèrdua d'informació en el sistema de preservació)*	Còpies en altres suports físics o en un CPD extern.	Actualment existeixen còpies en altres suports CD/DVD, discos externs. Existeix la voluntat de crear un mirror de les dades en un sistema similar en instal·lacions externes a la BC	

8.2.3 Gestió d'usuaris i seguretat

Es defineixen cinc perfils d'usuaris per al programari intern de gestió del repositori. Cada nivell superior acumula els drets dels inferiors. Exceptuant el dret de lectura, els drets es donen sempre dins d'un contenidor específic i s'hereten en el contenidors fills.

- Usuari bàsic: pot consultar el repositori, ho són totes les persones que treballen a la BC.
- Dipositant: pot crear documents, versions i penjar fitxers. No poden modificar dades, ni posar els documents en estat "publicat" que permetria la seva visualització externa. Ho són els editors/productors autoritzats de Dipòsit Legal.
- Editors: poden editar documents, versions i fitxers. Poden posar els documents en estat "publicat" per a permetre la seva visualització externa. Ho són els tècnics de digitalització de la BC.
- Revisor: poden posar en estat de "supressió" documents, versions i eliminar fitxers.
- Administrador de contenidor: pot crear i modificar contenidors fills, eliminar-los si són buits i assignar drets a usuaris per a un contenidor. Té accés al llistat de documents i versions en estat "suprimit" per recuperar-los o eliminar-los definitivament. Es defineix un administrador per a cada contenidor de primer nivell.
- Superadministrador: pot eliminar contenidors de primer nivell. Només l'administrador del sistema.

Es guarda un registre (logs) de totes les accions que es realitzen i s'establiran mecanismes de comprovació periòdica que permetin identificar si s'estan produint incidències o anomalies no desitjades (p.e. eliminació reiterada de fitxers).

8.2.4 Gestió de continguts: càrregues

Per als processos de creació de documents/versions manuals, el sistema permet la càrrega de fitxers individuals o de tots els fitxers continguts en una carpeta.
Es crearà un mòdul de plantilles que permeti establir metadades descriptives i tècniques per defecte en la càrrega de documents. Aquestes plantilles es podran seleccionar per contenidor.

Per a les càrregues manuals més complexes de documents que comprenen una jerarquia de carpetes s'establirà un perfil de càrrega que permeti la creació dels documents/versions i les càrregues de fitxers de manera semi/automàtica i que determinarà l'estructura de carpetes/documents en els futurs processos de digitalització tant interns com mitjançant concursos externs.

Així mateix, per a grans volums de documents amb estructures de carpetes diverses -i heterogènies entre sí en alguns casos- (que corresponen majorment a digitalitzacions ja existents), es generaran

càrregues en batch que s'adequaran a cada cas per a permetre una càrrega automatitzada.

8.2.5 Gestió del Dipòsit Legal de documents nascuts digitals

De la mateixa manera que els productors/editors de documents en suport físic poden realitzar la sol·licitud i gestió de números de DL via web, el sistema permetrà que dipositin el document en línia enlloc de lliurar-ne una còpia en DVD com es feia fins ara.

Els documents procedents de DL s'integraran al contenidor específic de DL, i se'ls aplicarà les mateixes rutines i accions de preservació que a la resta de documents.
Els editor/productors interaccionaran amb el sistema mitjançant una interfície web, seran tractats com a usuaris dipositants una vegada s'hagin identificat al sistema. S'estableixen com a pautes de dipòsit:

- Format dels fitxers: PDF preferentment
- Nom del fitxer: Núm. de dipòsit legal del document

Una vegada dipositat el document:

- El sistema genera un avís de correu electrònic a l'oficina de DL per a que validi el document
- L'oficina de DL si és correcte el valida, genera el formulari estàndard M5 (de dades descriptives del document) que obté del programa de DL del que ja disposa la BC i l'empaqueta amb el document en el repositori; si no és correcte, es generarà un avís a l'editor/proveïdor.
- Una vegada empaquetat el document i l'M5, s´enviarà per ftp a les biblioteques dipositàries corresponents, és a dir, Biblioteca Nacional d'Espanya, Biblioteques Públiques de Girona, Tarragona i Lleida.
- A la vegada s'envia un correu al servei corresponent de la Biblioteca de Catalunya per a que cataloguin el document, de manera que pugui ser consultat dins de la xarxa interna de la BC.

8.2.6 Gestió de còpies d'alta qualitat

La gestió de còpies d'alta qualitat per a ús comercial i d'investigació és un mòdul en entorn web que es desenvoluparà una vegada la gestió interna i de dipòsit legal del repositori de preservació estigui en producció.

Es tracta d'un mòdul complex que no s'ha finalitzat de dissenyar a nivell teòric i és prematur definir-ne les seves característiques en aquest informe.

8.2.7 Caixa negra

Com a sistema de preservació cal restringir al màxim l'accés als volums de dades, sense impossibilitar els objectius de servei designats en el punt 2. Usos, d'aquest informe.

S'estableixen dos mètodes d'accés a les dades:

- *Directe*
  Amb el mòdul de gestió interna del Repositori amb control per IP i identificació per LDAP. Limitat únicament al personal de la BC per a les càrregues i les còpies d'alta qualitat.
- *Indirecte*
  A través de serveis web obert al públic: accés des de catàleg, consulta directa a través de metadades, etc. amb drets únicament de lectura.

9 CALENDARI

Cal anotar que l'equip de desenvolupament actual de la BC és reduït i que la dedicació a aquest projecte no és en cap cas a temps complert. Per tant aquestes dates poden variar lleugerament degut a canvis en la planificació de prioritats a curt i mig termini.

	En producció	En desenvolupament	Planificat
Base de dades		Sí	
Checksum	Sí		
Antivirus	Sí		
Detecció de format per DROID	Sí		
Detecció de format per Libmagic	Sí		
Balanceig de càrrega en els volums de dades	Sí		
Càrrega de continguts manual	Sí		
Mòdul de plantilles de metadades		Sí	
Càrrega de continguts manual complexes			Gener 2011
Càrrega de continguts en batch			Gener 2011
Gestió d'objectes en estat suprimit	Sí		
Creació de logs	Sí		
Gestió d'usuaris i control d'accés	Sí		
Gestió de metadades	Sí		
Gestió de DL			Febrer 2011
Gestió de les rutines de preservació			Febrer 2011
Gestió de les còpies d'alta qualitat			Maig 2011
Serveis web de consulta pública			Maig 2011