

# Com guardar el món digital dins d'un COFRE (COservem per al Futur Recursos Electrònics)

**Karibel Pérez Villalba i Eugènia Serra Aranda**

Biblioteca de Catalunya

*kperez@bnc.cat, eserra@bnc.cat*

## **Resum:**

La preservació digital ha esdevingut una de les àrees d'interès i un repte per a les institucions que conserven patrimoni. Arreu del món, les institucions, sovint de caràcter nacional, han posat en marxa iniciatives per fer-hi front i explorar-ne solucions. Al nostre país, la Biblioteca de Catalunya, l'any 2008, va iniciar el disseny d'un repositori digital de preservació anomenat COFRE (COservem per al Futur Recursos Electrònics), que des de 2010 es troba operatiu. COFRE inicialment va ser impulsat per cobrir les necessitats pròpies de la Biblioteca, si bé sempre ha tingut la voluntat i la visió de poder acollir altres institucions i oferir així un servei de país. L'article en descriu l'origen, el contingut, els processos, les característiques tècniques i la infraestructura.

## **Paraules clau:**

preservació digital, Biblioteca de Catalunya, COFRE (COservem per al Futur Recursos Electrònics), repositoris digitals, biblioteques nacionals

## Cómo guardar el mundo digital en un COFRE (COservamos para el Futuro Recursos Electrónicos)

### Resumen:

La preservación digital se ha convertido en una de las áreas de interés y en un reto para las instituciones que conservan patrimonio. En todo el mundo, las instituciones, a menudo de carácter nacional, han puesto en marcha iniciativas para hacerle frente y explorar soluciones. En nuestro país, la Biblioteca de Catalunya, en 2008, inició el diseño de un repositorio digital de preservación: COFRE (COservamos para el Futuro Recursos Electrónicos), que desde 2010 se encuentra operativo. COFRE inicialmente fue impulsado para cubrir las necesidades propias de la Biblioteca, si bien siempre ha tenido la voluntad y la visión de acoger a otras instituciones y ofrecer así un servicio de país. El artículo describe su origen, el contenido, los procesos, las características técnicas y la infraestructura.

### Palabras clave:

preservación digital, Biblioteca de Catalunya, COFRE (COservamos para el Futuro Recursos Electrónicos), repositorios digitales, bibliotecas nacionales

## How to keep the digital world in a COFRE (COservamos para el Futuro Recursos Electrónicos)

### Abstract:

Digital preservation has become an area of interest and a challenge for the institutions dedicated to preserving heritage. Worldwide, institutions, often national ones, have launched initiatives to address the issue and to explore solutions. In Catalonia, the Biblioteca de Catalunya began designing a digital preservation repository known as COFRE (*COservem per al Futur Recursos Electrònics* = Preserving electronic resources for the future) in 2008. It has been in operation since 2010. COFRE was initially promoted to meet the needs of the library itself, though it has always held the intention and vision to house other institutions, thereby providing a service for the country. The article describes the background, content, processes, technical features and infrastructure of the repository.

### Key words:

digital preservation, Biblioteca de Catalunya, COFRE, digital repositories, national libraries

## 1. Introducció

La Biblioteca de Catalunya (BC) té entre les seves funcions preservar la producció bibliogràfica de Catalunya i en català. Aquesta funció, fins al darrer quart del segle passat, responia a un món analògic en el qual els documents es trobaven principalment sobre paper, i sobre el qual les biblioteques disposaven de coneixement, experiència i una llarga tradició de conservació i preservació.

El segle XXI ha suposat la generalització de l'ús d'Internet com a mitjà de publicació, comunicació i edició, fet que ha generat un nou model de producció «bibliogràfica» canviant i divers. Aquest context ha fet que les biblioteques i els centres de conservació hagin posat en marxa iniciatives per preservar, a més de l'analògic, el patrimoni digital, conscients que anirà creixent i evolucionant cap a tipus, formes i formats de documents que, probablement, avui en dia encara desconexem. Treballem, doncs, sota un prisma de necessària adaptació amb el pas dels anys.

## 2. Qui està preservant?

La preservació digital afecta tota mena de documents digitals, els que tradicionalment es conservaven a les biblioteques però també la documentació característica dels arxius; afecta la pervivència de l'herència cultural dels països, però també les dades generades en l'entorn de la recerca; és, en conseqüència, una inquietud col·lectiva d'ampli abast. Per donar-hi resposta, s'estan impulsant arreu del món diverses iniciatives i projectes, principalment

des d'institucions acadèmiques i nacionals, donat que la preservació digital requereix d'instal·lacions, procediments i infraestructures complexes amb un cost elevat no assequible per a qualsevol organització.

Les biblioteques nacionals són en molts casos les que han posat en marxa repositoris de preservació digital nacionals. Biblioteques referencials d'arreu del món tenen les seves pròpies iniciatives en curs, com ara el National Digital Information Infrastructure and Preservation Program<sup>1</sup> de la Library of Congress, o els diversos programes<sup>2</sup> de la British Library, als quals s'han sumat altres projectes, entre els quals, per la seva completesa i naturalesa ressenyem, a títol il·lustratiu, el National Digital Heritage Archive, que la National Library of New Zealand (*ndha-wiki.natlib.govt.nz/ndha/pages/NationalDigitalHeritageArchive*), en partenariat amb Ex Libris Group –un proveïdor de sistemes de gestió de biblioteques– i Sun Microsystems –proveïdor de sistemes oberts de computació en xarxa–, ha desenvolupat des de 2004. Consisteix en un sistema de preservació digital per a documents nascuts digitals i digitalitzats, que una vegada implementat a la National Library of New Zealand ha esdevingut comercial. El maquinari i el programari estan dissenyats per ser escalables en el temps a mesura que creixen les col·leccions. El sistema disposa d'eines per integrar-se amb el sistema de gestió bibliotecària i també d'eines de dipòsit en línia per als creadors/editors. Conté aproximadament 500.000 entitats intel·lectuals.

1. En el marc del National Digital Information Infrastructure and Preservation Program ([www.digitalpreservation.gov/about/index.html](http://www.digitalpreservation.gov/about/index.html)), la Library of Congress l'any 2010 va posar en marxa la National Digital Stewardship Alliance, una acció basada en la cooperació entre el sector públic i privat per a la creació d'eines, estàndards, recomanacions, etc., que contribueixin a garantir la preservació a llarg termini.
2. La British Library disposa de diferents programes relatius a la preservació de documents nascuts digitals i digitalitzats d'analògics tant per a textuais com per a sonors i audiovisuals. A la mateixa web de la British Library ([www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/index.html](http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/index.html)) es pot trobar documentació al respecte.

### 3. COFRE (COnservem per al Futur Recursos Electrònics), la iniciativa de la Biblioteca de Catalunya

#### 3.1 Objectius

El repositori de preservació digital COFRE té com a objectiu garantir la pervivència dels documents digitals fruit de la digitalització d'originals analògics i dels documents nascuts digitals, a Internet o en altres suports.

#### 3.2 Context

La necessitat de la creació d'un repositori de preservació sorgeix per a la BC vinculada a diversos fets:

a) La digitalització intensiva de documents en domini públic realitzada per difondre a Internet el patrimoni català a partir de 2006 fa que es generin moltes còpies digitals, còpies que, des del començament, s'han obtingut en alta qualitat per preservar, i a partir de les quals se n'obtenen les còpies menys pesants per publicar a Internet. Aquestes còpies d'alta qualitat es trobaven en diferents suports, principalment DVD i discs durs externs, fet que generava incertesa quant a la seva disponibilitat al cap dels anys

per degradació del suport. Es tractava de disposar d'un màster del qual obtenir-ne posteriors còpies en cas de petició dels usuaris o d'incidències de la versió publicada en obert. En aquest apartat s'inclouen les digitalitzacions fetes per ARCA (Arxiu de Revistes Catalanes Antiques)<sup>3</sup>, MDC (Memòria Digital de Catalunya)<sup>4</sup> i Google Llibres<sup>5</sup>.

- b) Han passat més de 30 anys des que la BC, per dipòsit legal, comencés a rebre CD textuals, sonsors i audiovisuals que presenten risc de degradació del suport, de pèrdua d'integritat de les dades, o que les dades estiguin generades en entorns de maquinari i programari obsolets, difícils de reproduir.
- c) La nova llei de dipòsit legal de juliol de 2011<sup>6</sup> i el real decret en procés de redacció<sup>7</sup> que regularà el dipòsit legal de les publicacions electròniques contempla en la captura o lliurament dels recursos nascuts digitals als centres que en faran la conservació, com és el cas de la BC a Catalunya. En conseqüència, calia crear un entorn en línia de lliurament de tots aquests tipus de documents per als creadors/editors que fan la comunicació pública oberta o restringida, i que estigui vinculat al sistema de preservació dels documents; es tracta de materials molt diversos, com ara música, llibre electrònic, diaris i revistes digitals, etc., que en el futur no tenim la certesa de les formes, formats i funcionalitats que presentaran.

Pel que fa a la preservació de la web de Catalunya, la Biblioteca de Catalunya recopila a través del sistema PADICAT

3. ARCA (Arxiu de Revistes Catalanes Antiques) [en línia]. Barcelona: Biblioteca de Catalunya, 2005- <<http://www.bnc.cat/digital/arca/index.html>> [Consulta: 24 de juny 2012]. Portal cooperatiu que conté la digitalització de 295 capçaleres catalanes que ja no es publiquen, que representen més d'1.200.000 pàgines accessibles en línia gratuïtament. És una iniciativa impulsada per la BC, de caràcter cooperatiu amb 22 institucions col·laboradores (biblioteques, arxius, museus, ateneus i associacions).
4. MDC (Memòria Digital de Catalunya) [en línia]. Barcelona: Biblioteca de Catalunya, Consorci de Biblioteques Universitàries de Catalunya, 2005- <<http://mdc.cbuc.cat>> [Consulta: 24 de juny 2012]. Portal cooperatiu impulsat conjuntament pel Consorci de Biblioteques Universitàries de Catalunya (CBUC) i la Biblioteca de Catalunya que compta amb més de 71 col·leccions de materials de tipologia diversa: cartells, llibres, manuscrits, mapes, fotografies, revistes, etc., de 21 institucions participants, que representen més de 2.000.000 d'imatges.
5. Google Llibres [en línia]. Google Inc. <<http://books.google.es/books?hl=ca>> [Consulta: 9 de juliol 2012]. L'any 2007 es feia públic el conveni signat entre Google Inc., la Biblioteca de Catalunya i quatre biblioteques catalanes més amb fons patrimonials (Biblioteca del Monestir de Montserrat, la Biblioteca Pública Episcopal del Seminari de Barcelona, la Biblioteca del Centre Excursionista de Catalunya i la Biblioteca de l'Ateneu Barcelonès), per adherir-se al projecte Google Llibres, digitalitzar de forma massiva llibres de domini públic (no subjectes a drets d'autor) i fer-los accessibles a través d'Internet. A 2012 pràcticament s'ha digitalitzat la totalitat de fons previstos, que representen més de 120.000 documents.
6. Llei 23/2011, de 29 de juliol, de dipòsit legal [en línia]. *Boletín Oficial del Estado, Suplement en llengua catalana* (30 de juliol de 2011), núm. 182, p. 1-12 <[http://www.boe.es/boe\\_catalan/dias/2011/07/30/pdfs/BOE-A-2011-13114-C.pdf](http://www.boe.es/boe_catalan/dias/2011/07/30/pdfs/BOE-A-2011-13114-C.pdf)> [Consulta: 24 juny 2012].
7. Els continguts del real decret s'estan treballant cooperativament en el Grupo de Trabajo de Depósito Legal del Consejo de Cooperación Bibliotecaria ([www.mcu.es/bibliotecas/MC/ConsejoCB/Presentacion.html](http://www.mcu.es/bibliotecas/MC/ConsejoCB/Presentacion.html)), amb representació de les diferents CCAA d'Espanya, i en el qual hi és present la BC.

(Patrimoni Digital de Catalunya)<sup>8</sup>, des del 2005, diverses versions de la web catalana. També d'aquest arxiu web se'n dipositaran a COFRE còpies anuals acumulatives.

### 3.3 Arquitectura actual

A nivell lògic global l'arquitectura està inspirada en el model OAIS (Open Archival Information System)<sup>9</sup>, marc de referència conceptual per a la preservació de la informació en format digital amb l'objectiu de garantir la seva accessibilitat en el futur. Aquest marc determina, a grans trets, les funcions, responsabilitats i organització d'un sistema de preservació. Aquest model, però, no especifica un pla d'implementació concret i per tant és obert a l'hora de definir el disseny del repositori de preservació.

A aquest model per a la part de preservació cal afegir les capes de servei específiques que la BC requereix. L'esquema complet del sistema queda recollit en l'esquema següent:

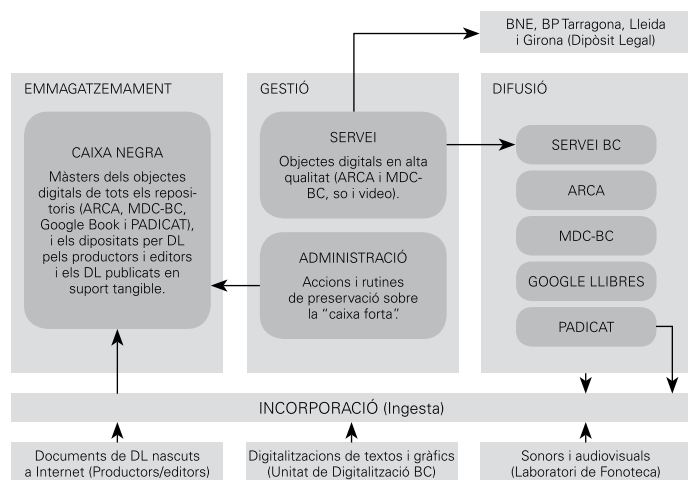


Fig. 1 Esquema de COFRE

#### 3.3.1 Infraestructura de la implementació

La implementació del repositori de preservació de la Biblioteca de Catalunya se sosté sobre la infraestructura següent:

- Base de dades que conté tota la informació del sistema:
  - Informació sobre l'estructura del sistema: esquemes de metadades emprats, jerarquia de contenidors, plantilles de metadades, usuaris, permisos i magatzems de dades.
  - Informació sobre els objectes emmagatzemats: informació descriptiva, administrativa, tècnica i de preservació dels objectes i els fitxers que els componen, la ubicació i l'estat a l'estructura de dades de l'aplicació i la ubicació en els magatzems de dades.
  - Informació de les accions sobre els objectes: *logs* de modificació, eliminació, etc., infeccions per virus, fitxers deteriorats o modificats.
- Magatzems de dades que emmagatzemen els fitxers digitals que componen els objectes. Per gestionar els magatzems s'ha desenvolupat un API que permet suportar diverses tecnologies d'accés a disc, com ara disc local, que és la que es fa servir actualment, o tecnologies d'accés per xarxa.
- Programari de càrrega, gestió i administració del repositori. Es va dissenyar la solució de manera que totes les accions a realitzar sobre el sistema, tant per part dels productors dels objectes digitals com dels clients potencials que han de poder consultar i extreure les dades, es fessin a través d'un client web. El programari web fa servir l'arquitectura *Model-Vista-Controlador*, que permet generar diverses vistes de les dades segons l'operació que s'ha de realitzar sobre aquestes, millorant així la reusabilitat del codi en el casos en què s'hagi de modificar la lògica de la interfície d'usuari.

8. PADICAT (Patrimoni Digital de Catalunya) [en línia]. Barcelona: Biblioteca de Catalunya, 2005- <[www.padicat.cat](http://www.padicat.cat)> [Consulta: 24 juny 2012]. A la mateixa web del sistema es troba bibliografia actualitzada al respecte, a més de la presentació, detalls tècnics i contingut.
9. Reference Model for an Open Archival Information System (OAIS) [en línia]. Consultative Committee for Space Data Systems, CCSDS 650.0-B-1, Blue Book, Issue 1, January 2002 <<http://public.ccsds.org/publications/archive/650x0b1s.pdf>> [Consulta: 24 juny 2012]; CCSDS 650.0-M-2: Magenta Book, Issue 2. June 2012 <<http://public.ccsds.org/publications/archive/650x0m2.pdf>> [Consulta: 24 juny 2012]. La darrera versió clarifica alguns conceptes i n'amplia d'altres respecte al document d'estàndards, si bé les noves especificacions no s'han contrastat amb les especificacions de COFRE.

- Per una altra banda, s'han desenvolupat *scripts* per a les rutines de preservació, que s'executen a nivell de sistema operatiu i que es realitzen cíclicament sobre els objectes digitals.

### 3.3.2 Arquitectura de l'aplicació

A nivell d'aplicació, l'arquitectura del sistema es pot dividir en tres grans mòduls, segons les seves funcionalitats:

- a) Mòduls de càrrega de continguts.  
Incorporen funcions de selecció dels objectes digitals ja sigui de manera individual, per a càrrega de multivolums o publicacions en sèrie, de comprovació de la seva validesa (integritat, infeccions per virus, format dels fitxers) i d'incorporació de metadades descriptives, administratives, tècniques i de preservació de l'objecte. Així mateix, incorpora les funcions relatives a la distribució òptima dels fitxers dels objectes en els magatzems de dades físics.
- b) Mòduls de gestió de dades.  
Incorporen funcions per al manteniment de metadades, tant descriptives com administratives, i el manteniment de plantilles de metadades per al seu ús en la càrrega de continguts.
- c) Mòduls d'administració.  
Incorporen funcions relacionades amb la seguretat de les dades i del sistema: creació de contenidors, assignació de permisos als usuaris, comprovació i recuperació dels elements eliminats, consulta del registre d'accions realitzades sobre els elements del sistema: creació, modificació, eliminació, etc. Inclou també les funcions de manteniment dels esquemes de metadades.

### 3.3.3 Estructura de les dades

L'estructura de les dades està suportada per cinc elements fonamentals: contenidors, documents, versions, fitxers i esdeveniments.

Els continguts del repositori s'organitzen per **contenidors**. Aquest element pot contenir altres contenidors i documents. La seguretat a nivell d'usuari és suportada per aquest element del sistema, de tal manera que els permisos assignats a un usuari per a un contenidor seran els que li donaran els drets sobre els objectes que conté. També és l'element al qual se li assignen **plantilles de metadades** com a valors per defecte en la creació de nous objectes.

El **document** és la entitat intel·lectual mínima. Per exemple, una monografia, un mapa, un número d'una revista, una entrevista radiofònica o una pel·lícula de vídeo. El document conté les metadades mínimes per a fer-ne la descripció bibliogràfica i els drets associats al document.

Cada document pot tenir una o més **versions**. Cada versió serà una representació completa del document, que pot variar en el format (tiff, jpeg, wav, mp3, etc.), segons hagi estat digitalitzat, canvis de versions, migracions a formats més estàndard, etc. Cada versió conté les metadades tècniques per descriure'l.

La versió conté un o més **fitxers**, sobre els quals es realitzen les accions de comprovació d'integritat i accessibilitat i se'n guarda el format, la mida i la data de càrrega.

Cada versió pot contenir un o més **esdeveniments** que descriuen accions realitzades sobre els fitxers de la versió i que corresponen a les metadades de preservació del document.

En el moment de la creació d'una versió per a la càrrega de fitxers, el sistema permet crear un esdeveniment «de creació», per tal de registrar l'acció que ha permès la creació d'aquests fitxers: digitalització, nascut digital, migració, etc.

## 3.4 Metadades de preservació i flux de treball

Per tal de poder assegurar l'accessibilitat dels objectes al llarg del temps el repositori emmagatzema dos tipus de metadades: les metadades tècniques i les metadades de preservació en format d'esdeveniments.

«L'estructura de les dades està suportada per cinc elements fonamentals: contenidors, documents, versions, fitxers i esdeveniments.»

### 3.4.1 Metadades tècniques

Informen del format dels fitxers i altres característiques associades a aquest format i segons el tipus d'arxiu (imatge, àudio, text, etc.). Aquestes metadades tècniques estan informades a nivell de versió del document i donen idea de les condicions tècniques en les quals es va digitalitzar o crear digital aquesta representació del document.

S'ha fet servir un esquema específic BC per a la descripció de les metadades, però s'ha basat la seva creació en esquemes actuals com MIX<sup>10</sup> i PBCORE<sup>11</sup>.

Fitxers	Format	PUID Pronom	Tipus MIME	Mida	Data càrrega
0913_28IV11BCIEC_0001.tif	TIFF image data, little-endian	fmt/10,fmt/7,fmt/8,fmt/9	image/tiff	38.99MB	07/06/2012
0913_28IV11BCIEC_0002.tif	TIFF image data, little-endian	fmt/10,fmt/7,fmt/8,fmt/9	image/tiff	38.48MB	07/06/2012
0913_28IV11BCIEC_0003.tif	TIFF image data, little-endian	fmt/10,fmt/7,fmt/8,fmt/9	image/tiff	38.79MB	07/06/2012
0913_28IV11BCIEC_0004.tif	TIFF image data, little-endian	fmt/10,fmt/7,fmt/8,fmt/9	image/tiff	38.06MB	07/06/2012

Fig. 2 Detall de la càrrega de fitxers

### 3.4.2 Metadades de preservació: esdeveniments

Per una altra banda, es poden establir les metadades de preservació en forma d'esdeveniments associats a cada versió del document. Aquestes metadades informen dels processos realitzats sobre la versió (digitalització, nascut digital, migració de format...), el productor que ho realitza i els dispositius usats. Aquests esdeveniments s'inspiren en el format PREMIS<sup>17</sup>.

Contestació a l'amic Sagimon		Codicillus seu vademecum	
Versió: <b>Alta qualitat</b>		Versió: <b>Imatges TIFF</b>	
<b>ESDEVENIMENTS</b>		<b>ESDEVENIMENTS</b>	
<input type="button" value="Editar"/> <input type="button" value="Eliminar"/>		<input type="button" value="Editar"/> <input type="button" value="Eliminar"/>	
Descripció	<b>Digitalització</b>	Descripció	<b>Digitalització</b>
Data	11/03/2008	Data	2009
Productor	Fonotron	Productor	Archygest
Dispositiu	Archeophone	Dispositiu	Metis DRS A1

Fig. 3 Exemples d'esdeveniments

### 3.5 Flux de treball

Donat que en aquests moments totes les càrregues de dades es realitzen a les instal·lacions de la BC, el flux de treball està molt simplificat. Està previst que la càrrega de documents nascuts digitals arribats a través del Dipòsit Legal i realitzada directament pels creadors/editors comporti un augment dels passos per tal de ser admesos en el repositori digital i que poden implicar diverses unitats de la BC.

- MIX (NISO Metadata for Images in XML): Technical Metadata for Digital Still Images Standards [en línia]. The Library of Congress <<http://www.loc.gov/standards/mix>> [Consulta: 24 de juny de 2012].
- PBCORE Public Broadcasting Metadata Dictionary Project [en línia] <<http://pbcore.org/index.php>> [Consulta: 24 de juny de 2012].
- LIBMAGIC [en línia] <<http://sourceforge.net/projects/libmagic>> [Consulta: 24 de juny de 2012].
- FIDO (Format Identification for Digital Objects) [en línia]. Open Planets Foundation <<http://www.openplanetsfoundation.org/software/fido>> [Consulta: 24 de juny de 2012].
- DROID (Digital Record Object Identification) [en línia]. The National Archives (UK) <<http://sourceforge.net/projects/droid>> [Consulta: 24 de juny de 2012].
- IANA (Internet Assigned Numbers Authority) [en línia]. The Internet Corporation for Assigned Names and Numbers <<http://www.iana.org/assignments/media-types/index.html>> [Consulta: 9 de juliol de 2012].
- PRONOM technical registry service [en línia]. The National Archives (UK) <<http://www.nationalarchives.gov.uk/PRONOM>> [Consulta: 24 de juny de 2012].
- PREMIS (Preservation Metadata: Implementation Strategies) [en línia]. The Library of Congress <<http://www.loc.gov/standards/premis>> [Consulta: 24 de juny de 2012].

La càrrega d'un document es pot fer de manera individual o en volum si tenen una estructura de carpetes nidada d'un nivell. Aquest segon mètode permet carregar diversos volums d'un mateix document o diversos números d'un títol d'una revista. En qualsevol dels dos casos (individual o en volum) el sistema obliga a seleccionar el contenidor on s'ubicaran els documents.

Una vegada seleccionat el contenidor, el sistema presenta la pantalla de camps del document on només la metadada «títol» és obligatòria. Si el document és al catàleg, el sistema proporciona una funció que permet copiar dades directament del catàleg bibliogràfic Millennium de la BC, en concret: títol, autor, número de dipòsit legal i data de publicació.

Si a més introduïm el número de registre d'exemplar del catàleg Millennium, el sistema ens permet copiar del catàleg els valors de topogràfic i codi de barres.

Fig. 4 Afegir document

Fig. 5 Afegir versió

En el moment de *Desar* el document podem també crear-ne una versió.

La pantalla de creació de la versió permet introduir una *Descripció* de la versió, que és obligatòria. Encara que aquest camp és d'entrada lliure, s'ha creat un desplegable amb les opcions més comunes. Així s'evita una bona part dels errors d'introducció de dades i es normalitza aquest valor.

La versió es pot marcar com a *Principal* o no. Aquest camp dona idea de quina de totes les versions d'un document és farà servir com a millor versió per al mòdul de gestió de còpies d'alta qualitat. El camp *Estat* també es fa servir per al mòdul de gestió de còpies d'alta qualitat. Només aquelles versions en estat «publicat» seran visibles des d'aquest mòdul.

Es mostren les metadades tècniques i es permet la introducció d'un esdeveniment de creació.



Metadades tècniques				
Text	Imatge	Audíio	Audiovisuals	Aplicació
Format: TIFF		Resolució: 400 ppp		Profunditat de bits: 24 bits
Canals: 3				

Fig. 6 Detall de la introducció de metadades tècniques

Per últim, existeixen dues opcions per a la càrrega de fitxers:

Si ocupen menys de 2GB en total, es pot fer una càrrega directa des de qualsevol ubicació local dels fitxers: disc de l'estació de treball, disc extern, memòria USB, etc.

Si ocupen més de 2GB en total, per qüestions de limitació del servidor web (Apache) cal primer copiar-los en un espai intermedi i informar al sistema de la ubicació dels fitxers. Aquesta última opció només és possible quan les càrregues es realitzen des d'una ubicació interna de la BC, i no es donarà com a opció per a les càrregues externes del Dipòsit Legal.

Durant el procés de càrrega el sistema fa servir diferents eines per reconèixer el format dels fitxers, calcular la signatura digital en format MD5 de cadascun d'ells i comprovar les infeccions per virus. Per a grans càrregues la comprovació d'infeccions per virus es fa a través d'*scripts* nocturns programats de manera cíclica.

Una vegada carregats els fitxers el document ja estarà disponible per a la seva visualització en el sistema de preservació. No requereixen, per tant, de passos suplementaris per a la seva aprovació. El sistema genera avisos d'infeccions per virus o errors d'integritat de fitxers per tal que els administradors del repositori comprovin i corregeixin els errors.

Depenent dels permisos d'usuari, a partir de la càrrega d'un document es podran afegir altres versions i esdeveniments, modificar les metadades, assignar-ho a un altre contenidor o eliminar elements a nivell de document, versió, esdeveniment o fitxer. Totes aquestes accions són registrades a la base de dades per a la seva posterior consulta.



©istockphoto/ Lise Gagne

En els casos d'eliminació de documents o versions, aquestes passen a estat «suprimit» i són els administradors del repositori qui validen l'eliminació o restauren el document.

Recuperar documents i/o versions suprimides	
Versions suprimides de documents no suprimits	Marca tot   Desmarca tot
<b>Documents suprimits</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> Paraula de poeta [Proves]           <ul style="list-style-type: none"> <li>Versions del document               <ul style="list-style-type: none"> <li><input type="checkbox"/> Digitalització JPEG</li> </ul> </li> </ul> </li> </ul>	
<b>Versions suprimides de documents no suprimits</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> El Alud verde [Proves]           <ul style="list-style-type: none"> <li>Versions del document               <ul style="list-style-type: none"> <li><input type="checkbox"/> Digitalització WAV</li> </ul> </li> </ul> </li> </ul>	
Documents suprimits	Recuperar

Fig. 7 Pantalla per a la recuperació de documents i versions suprimides

### 3.6 Desenvolupaments i millores futures

El repositori disposa actualment de les funcionalitats nuclears i bàsiques de preservació, a les quals s'afegiran a curt termini noves prestacions.

Vinculat estrictament a la preservació digital s'hi desenvoluparà:

- El mòdul d'anàlisi de riscos i de generació de plans de preservació específics que actuïn sobre les dades dipositades i que donin avís de les accions requerides.
- Capa de càrrega en línia de documents de DL per part dels creadors/editors.
- Millores en el control d'eliminació de documents. Mostrar dades en les pantalles d'eliminació i recuperació: usuari, data i hora, enllaç als *logs* de l'usuari.
- Llistats de control dels continguts: objectes sense registre bibliogràfic informat, documents amb versions sense fitxers, revisions de topogràfics i bibliogràfics contra catàleg.
- Canvis sobre metadades en conjunts d'objectes pre-seleccionats.

Vinculat a serveis addicionals del repositori s'hi desenvoluparà:

- Capa d'autoservei de còpies d'alta qualitat.
- *Scripts* anuals d'actualització dels drets de l'objecte (accés públic o restringit) segons la data dels drets especificada.
- Multiselecció de fitxers per a descàrregues internes.
- Exportacions de metadades METS.
- Servei OAI.

Un altre aspecte que resta pendent per a 2013 són les còpies de seguretat. Actualment el COFRE no disposa d'una còpia de seguretat en cinta o en disc que permeti mantenir una còpia completa dels volums de dades del repositori i la base de dades del sistema.

L'objectiu de la còpia de seguretat externa és tenir una segona còpia de les dades que ens permeti recuperar les dades en cas de pèrdua o desastre. No s'ha previst en principi donar servei d'alta disponibilitat; només permetre recuperar aquells fitxers malmesos o perduts, en cas necessari.

En aquests moments s'estan mantenint converses amb diversos proveïdors per tal d'estudiar diverses solucions

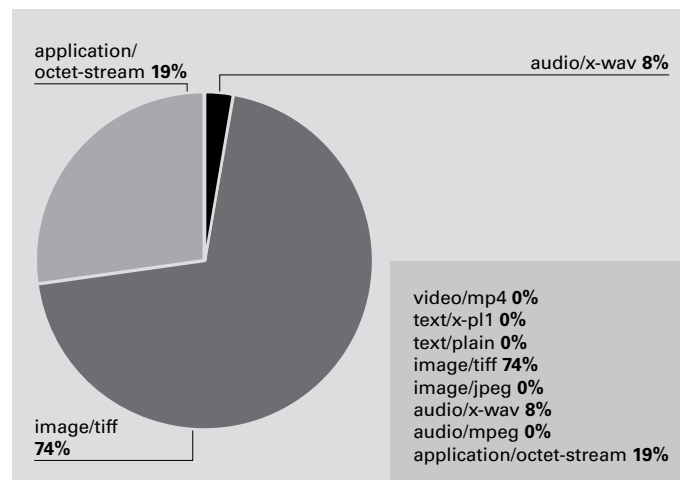
per realitzar aquesta còpia. Unes aposten per realitzar aquesta còpia al núvol, entenent-lo com un núvol privat d'accés restringit.

### 3.7 Contingut i previsió de creixement per als propers anys

A data juny de 2012, COFRE conté les següents quantitats totals:

- 82 contenidors
- 51.572 documents
- 51.705 versions
- 13.842.482 fitxers que ocupen un total de 8,27TB

La següent il·lustració mostra l'espai ocupat segons el tipus MIME del fitxer:



**Fig. 8** Contingut segons tipus de fitxers

El creixement de COFRE depèn de tres factors, essencialment, que vénen profundament limitats pel context actual de crisi pressupostària: d'una part, dels plans de digitalització que puguin dur a terme la BC i que, per tant, proveiran de major o menor nombre d'objectes digitals a carregar; de l'altra, del recursos humans disponibles per realitzar les

càrregues, la introducció de metadades i el seguiment del procés; i per últim, de la possibilitat d'adquirir una ampliació de disc per ampliar els volum de dades.

Tenint en compte els objectes digitals existents en l'actualitat, les previsions de creixement de continguts de la pròpia BC són, per al 2012-2013:

- Àudio i vídeo: 3TB
- Imatge: la quantitat existent d'objectes digitals supera la capacitat actual de COFRE, per tant, en les actuals condicions, es podrien carregar un total de 8TB.

COFRE va néixer com a sistema de preservació digital de la Biblioteca de Catalunya, però des de l'inici, i encara més a mesura que s'avançava en la definició de requeriments, primer, i en el desenvolupament, després, va prendre força la idea que havia d'estendre's a altres institucions i constituir-ne un sistema nacional.

Adquirir o desenvolupar sistemes de preservació digital és costós, tant en termes d'infraestructura de maquinari com d'adquisició de coneixement especialitzat i d'experiència. És, doncs, una iniciativa que no poden abordar moltes institucions del país, que sí, en canvi, estan implicades en la generació de continguts digitals, ja sigui per digitalització com produïts per la pròpia organització. Els avantatges de treballar en cooperació són evidents i reconeguts en qualsevol àmbit, i en un sistema com és COFRE es fan encara més evidents: l'economia d'escala, l'aprofitament de coneixement, la garantia de permanència de la institució que l'impulsa són, entre d'altres, bons arguments per constituir-ne un repositori nacional.

En aquesta línia de treball, el 2012 la Biblioteca de Catalunya ha signat un conveni amb l'Ateneu Barcelonès perquè pugui preservar a COFRE les seves còpies procedents de la digitalització. S'ha replicat una estructura de contenidors específica amb la seva pròpia definició d'usuaris i permisos per donar-hi cabuda. L'Ateneu Barcelonès contribueix econòmicament al manteniment del sistema, proporcionant a l'espai que ocupa.

L'ocupació actual a COFRE de l'Ateneu Barcelonès és la següent:

- 11 contenidors
- 4.966 documents
- 4.966 versions
- 2.196.314 fitxers que ocupen un total de 839,7GB

Les previsions d'ocupació són de 3,5TB el 2012, 3TB el 2013 i 2TB el 2014.

L'experiència de la participació de l'Ateneu proporcionarà una informació molt valuosa en termes d'avaluació del sistema, que permetrà plantejar-ne altres incorporacions futures.

### 3.8 Servei de còpia i oportunitats al núvol

En aquests moments estem fent un estudi de les possibilitats respecte quina solució adoptar per realitzar una segona còpia del repositori de preservació. Les qüestions plantejades fins ara són, entre d'altres, si la còpia s'ha de mantenir a les instal·lacions de la BC o fora d'elles, quina hauria de ser la vida de la còpia (donat el gran volum de dades existent, podria passar que es perdés o es fes malbé un fitxer i que no es detectés fins passat un temps considerable), si cal que la còpia sigui síncrona o pot ser una còpia de dades amb una finestra temporal limitada, si la còpia ens ha de proporcionar alta disponibilitat del sistema de preservació i, finalment, si es pot adquirir la solució de còpia com un servei de tercers.

Inicialment es va plantejar amb força l'adquisició d'un segon magatzem físic de dades, a la BC o a les instal·lacions d'un proveïdor de serveis d'allotjament. Es va pensar en sistemes amb discos d'accés més lents però de gran capacitat, el cost dels quals fos més assequible i que incorporessin funcions d'eficiència energètica. Però donat l'increment actual de serveis d'emmagatzemament basats en cost per GB es requereix un replantejament de la solució.



©istockphoto/Alex Slobodkin

Aquests serveis són, en alguns casos, solucions d'emmagatzemament al núvol, i d'altres, d'emmagatzemament privat.

Els avantatges de la còpia remota com a servei són importants:

- No existeix una forta despesa inicial de maquinari i programari. Es paga per ús.
- Permet l'administració pròpia de les còpies: polítiques de les còpies, recuperació de dades, etc., com si el servei fos en local.
- Assegura un cert control d'accés a les dades.
- Es té accés a recursos tecnològics superiors sense augment, necessàriament, dels costos del servei.
- Permet un creixement progressiu de l'emmagatzemament sota petició, sense temps d'espera per a l'adquisició de discos nous, instal·lació, etc.
- Allibera el personal tècnic intern de la gestió d'un altre magatzem de dades i allibera dels costos de manteniment i infraestructura associats.

Però queden pendents dubtes importants. En el cas de solucions en el núvol, quins convenis sobre control

d'accés o seguretat de les dades, sobretot en el cas de documents amb drets de propietat intel·lectual, ha subscrit el proveïdor del servei amb els proveïdors finals de l'emmagatzemament? Si els proveïdors de l'emmagatzemament es troben a altres països, de quins termes legals i de privacitat disposes? Com podem assegurar o quin nivell de seguretat tenen les nostres dades si han estat «barrejades» amb les d'altres clients, amb tècniques de duplicació i compactació, i qui té finalment accés a elles des del propi proveïdor del servei? A més, creiem que és important definir un pla de no continuació del servei: com es recuperarien aquestes còpies en el cas de trencament de contracte, de dissolució o canvi de model del servei? Es podrien migrar a un altre proveïdor de serveis o als models de servei disponibles en un futur? Quin és el temps garantit de funcionament del servei?

Actualment, per tant, estem en fase de prospecció i mantenint converses amb diversos proveïdors per tal d'estudiar les solucions possibles per realitzar aquesta còpia.

## 3.9 Cost del projecte i vies de finançament

### 3.9.1 Costos del maquinari

Com a costos de maquinari es tenen en compte els costos per adquirir el volum de dades i el servidor que gestiona el programari, el servidor web i la base de dades.

La inversió en maquinari va ascendir a 122.551,67 € entre els anys 2008-2010.

### 3.9.2 Costos en recursos humans

El disseny i desenvolupament de COFRE s'ha fet, exclusivament, amb tècnics de la plantilla de la BC. La dedicació en hores ha estat:

- Grup de treball de definició de requeriments: 100 hores (còmput total del treball individual i col·lectiu)
- Desenvolupament de la base de dades i de les funcionalitats bàsiques de preservació:
  - 600 hores de programador sènior.
  - 150 hores de disseny web.
  - 150 hores de cap de projecte.

El manteniment de l'aplicació requereix una dedicació mínima.

S'estima que el desenvolupament dels mòduls i de millores vinculades a preservació descrites en el punt 3.6 suposaran un 60% de la dedicació del desenvolupament inicial.

El calendari previst de finalització de tots els mòduls i millores és 2014, si bé resta supeditat a les necessitats de desenvolupament o gestió d'altres projectes o serveis existents de la BC.

La globalitat dels costos s'ha finançat amb el pressupost ordinari de la Biblioteca de Catalunya. No es contemplen en aquest apartat els costos corresponents a la càrrega dels objectes digitals en el repositori, que es fa de forma distribuïda entre diverses unitats de la biblioteca.

## 4. Conclusions

COFRE és necessari, està operatiu i és un eina que compleix els requisits necessaris per poder garantir la pervivència del patrimoni digital. COFRE, quan disposi dels mòduls en curs de desenvolupament, donarà una solució integral a la pervivència de les còpies, fruit de la digitalització, servirà de còpia de seguretat del sistema PADICAT i proporcionarà la tecnologia i l'eina de dipòsit per als creadors/editors. Resten, però, diversos reptes i decisions a prendre.

En l'àmbit de la gestió de la col·lecció de documents nascuts digitals, cal definir una política clara i alhora flexible de quines variants de documents nascuts digitals cal preservar; per exemple, si parlem de diaris digitals, què guardarem? Només la versió per a web? Les adaptades a dispositius mòbils? Per a quins? Totes les versions?

Pel que fa a la voluntat de donar un servei nacional, en els propers anys caldrà articular una estratègia de cooperació que resulti còmode i eficient per a les institucions participants i permeti que COFRE sigui un sistema escalable, sostenible i consolidat.

En definitiva, queden encara força incògnites, però la Biblioteca de Catalunya ha fet un primer pas, que entenem que forma part de la seva tasca com a biblioteca nacional i que pot, a la vegada, dibuixar una línia de treball col·lectiva per al país.