

Biblos: el projecte de conversió retrospectiva de la Biblioteca de Catalunya

Eugènia Serra

RESUM: *Exposició del projecte Biblos per a la conversió retrospectiva dels catàlegs manuals de la Biblioteca de Catalunya, que consisteix en l'escaneig de les fitxes del catàlegs, la recuperació de les imatges via web, i la posterior codificació de la informació en format MARC. Se'n descriuen els objectius, la metodologia, els procediments i els terminis previstos per a la seva realització.*

La Biblioteca de Catalunya (d'ara en endavant BC), el mes de setembre de 1997 posà en marxa el projecte Biblos per a la conversió retrospectiva dels seus catàlegs manuals. El projecte es definí conjuntament amb l'empresa VTLS, que és qui l'executa, amb el suport del personal i de les infraestructures de la biblioteca. En aquest article n'exposarem els objectius, la metodologia emprada, en què consisteix i com ha estat dissenyat.

1. ANTECEDENTS

La BC fou fundada el 1907 com a biblioteca de l'Institut d'Estudis Catalans i, el 1914, la Mancomunitat de Catalunya la convertí en una biblioteca de recerca i de conservació del patrimoni bibliogràfic català. Des dels seus inicis fins al 1990 els processos de gestió documental es realitzaren de forma manual.

L'any 1991 s'inicià la informatització de la catalogació corrent dels documents amb l'eina informàtica SICAB,¹ desenvolupada i gestionada pel Centre

Eugènia Serra. Biblioteca de Catalunya
Article rebut el febrer de 1998

1. Ramon, Mercè; Lagüens, Assumpta, "Catàleg automatitzat de la Biblioteca de Catalunya", en: *4es Jornades Catalanes de Documentació: biblioteques, centres de documentació i serveis d'informació* (Barcelona: Societat Catalana de Documentació i Informació: Col·legi Oficial de Bibliotecaris-Documents de Catalunya, 1991)

Informàtic de la Generalitat de Catalunya, que cobria les necessitats de la BC quant a la catalogació dels documents i a la consulta OPAC; posteriorment, el Centre Informàtic desenvolupà un programari per a la gestió de les adquisicions, que no era, però, integrat amb la catalogació ni amb la consulta. El 1995 la BC es replantejà la seva informatització i decidí d'adquirir un paquet comercial que cobrés tota la gestió del procés documental, des de les adquisicions fins a la circulació de documents. Amb aquesta finalitat obrí un concurs públic, al qual optaren diverses empreses, el resultat del qual fou la selecció del sistema VTLS, el mateix que prèviament havien seleccionat totes les universitats públiques catalanes i la Diputació de Barcelona.

El mes de desembre de 1995 la BC obrí al públic l'OPAC amb aquest nou sistema i hi inicià la catalogació corrent. Al llarg de 1996 s'implementaren progressivament els altres subsistemes: circulació de documents, control de sèries i adquisicions.

Posteriorment, el primer trimestre de 1997, s'implementà una passarel·la per a accedir a la base de dades de la biblioteca a través d'Internet, via web.² Aquesta situació cobria els objectius bàsics de la biblioteca quant a la difusió dels documents ingressats a partir de 1991; quedava, però, per solucionar la difusió i accessibilitat al fons documental catalogat abans d'aquesta data, que representa un volum aproximat de 800.000 documents, davant dels 225.000³ que integren la base de dades en línia; calia dissenyar un projecte per la conversió retrospectiva dels catàlegs.

2. CARACTERÍSTIQUES DELS CATÀLEGS MANUALS

La BC disposa de dos catàlegs manuals alfabètics d'autors i dos catàlegs manuals alfabètics de matèries:

- Catàleg alfabètic d'autors i obres anònimes: conté, aproximadament, 800.000 fitxes, principals, secundàries, de referència i analítiques, dels documents ingressats per compra, donatiu o intercanvi a la BC des de la seva creació fins a l'any 1990. El format de les fitxes és de cedulari i el tant per cent de fitxes redactades a mà és del 55%.

2. Biblioteca de Catalunya, *Catàleg en línia* <<http://www.genocat.es/bc/virtua/catalan/>>.

3. Aquesta xifra és el resultat de la unió en una sola base de dades, l'any 1995, en canviar de sistema informàtic, dels registres bibliogràfics de l'Hemeroteca Nacional de Catalunya, de la Fonoteca Nacional de Catalunya, de la Biblioteca Bergnes de las Casas –institucions que d'acord amb la llei havien quedat integrades dins de la BC– i de la Biblioteca de Catalunya.

- Catàleg alfabètic d'autors-Dipòsit legal: conté, aproximadament, 150.000 fitxes, principals, secundàries i de referència, dels documents arribats per dipòsit legal des de 1982 fins a 1990. El format de les fitxes és estàndard (7,5 x 12,5 cm) i són fetes en un 100% a màquina.
- Catàleg alfabètic de matèries en castellà: conté, aproximadament, 235.000 fitxes, els encapçalaments de les quals són redactats en castellà, i corresponen als documents catalogats fins al 1981.⁴ El format de les fitxes és estàndard i el percentatge de fitxes a mà és del 55%.
- Catàleg alfabètic de matèries en català: conté, aproximadament, 115.000 fitxes, els encapçalaments de les quals són redactats en català, i corresponen als documents catalogats des del 1982 fins al 1990. El format de les fitxes és estàndard i el percentatge de fitxes a mà és del 25%.

En resum, les fitxes dels catàlegs de la BC han estat realitzades en formats diferents: mida de cedulari i mida estàndard, moltes han estat redactades a mà (45%), amb nombroses grafies diferents, d'altres són fetes a màquina (55%), i les normatives per a redactar-les han canviat al llarg dels anys, des de les *Instrucciones*⁵ fins a les ISBD.

3. ELECCIÓ DEL SISTEMA DE RECONVERSIÓ

Per a procedir a la conversió retrospectiva dels catàlegs la Biblioteca va recollir informació dels resultats i dels costos dels diferents sistemes de reconversió utilitzats en el seu entorn més pròxim: conversió a partir de fotocòpies de les fitxes, catalogació per còpia d'altres bases de dades en línia, contractació de personal temporal..., a través de diversos contactes amb les institucions i/o les empreses implicades i fins i tot es varen realitzar algunes proves de conversió amb una selecció de fitxes, en base a la catalogació a partir de fotocòpies. Posteriorment, assabentats per la pròpia BC de la voluntat de posar en marxa aquest projecte, VTLS oferí a la biblioteca una solució força interessant i innovadora en el nostre país: es tractava de digitalitzar les fitxes dels catàlegs

4. La proporció entre un i altre catàleg de matèries respon al fet que la redacció de fitxes de matèries no començà fins a la dècada dels trenta. En aquells moments el catàleg s'incia en català, però després de la guerra fou castellanitzat.
5. *Instrucciones para la redacción del catálogo alfabético de autores y obras anónimas en las bibliotecas públicas del Estado*. 2a. ed. (Madrid: Dirección General de Archivos y Bibliotecas, 1941) (altres edicions: 1955, 1960, 1964, 1970). Anteriorment a les *Instrucciones*, adoptades als anys quaranta, la Biblioteca de Catalunya utilitzava unes regles pròpies.

manuals, les imatges resultants fer-les accessibles per Internet i codificar la informació posteriorment a partir d'aquestes imatges. L'avaluació comparativa de les diferents opcions recomanà com a solució més satisfactòria la proposta per aquesta empresa.

L'estiu del 1996, la BC lliurà a l'empresa un informe quantitatiu i descriptiu dels catàlegs, per a poder realitzar un estudi que permetés definir resultats, costos i terminis, amb més exactitud. El mes de setembre de 1996 es va fer un estudi previ en base a l'opinió d'un expert nord-americà que va analitzar el material a reconvertir i la problemàtica que podia representar. A finals de 1996, la BC disposava del pressupost de dues possibles solucions:

- Escanneig de les fitxes i creació d'una base d'imatges consultable via Internet mitjançant índexs html: oferia la possibilitat d'obtenir resultats en poc temps –el termini d'execució previst era de tres mesos–, el cost assequible i permetia donar a conèixer les col·leccions de la BC a través d'un sistema de consulta senzill, com és la consulta via web. Aquesta solució, però, no cobria totes les necessitats de la BC que, com a biblioteca nacional, ha de disposar d'una base de dades única, amb registres codificats en format llegible per màquina que permeti la importació i exportació de registres amb altres institucions.
- Escanneig de les fitxes només amb la finalitat de la seva conversió a format MARC i, per tant, sense la consulta en línia de les imatges: proporcionava una base de dades amb registres codificats en format estàndard, el temps d'execució era més llarg, el cost, lògicament, més elevat i requeria igualment realitzar l'escanneig de les fitxes per a disposar de la informació que calia codificar.

Una vegada analitzades ambdues propostes, la BC va resoldre que calia desenvolupar una solució mixta: creació d'una base de dades d'imatges consultable per Internet a través del web, en una primera fase, i conversió manual de les dades de les fitxes a registres codificats, a partir de les imatges i amb l'ajuda d'unes plantilles predefinides per la biblioteca, en una segona fase.

La decisió final de seleccionar VTLS com a proveïdor per a la reconversió es basà en els següents criteris:

- experiència de l'empresa en la digitalització de documents: la BC va mantenir diversos contactes amb les institucions per a les quals ja havia realitzat projectes similars de digitalització per tal de conèixer-ne els resultats, la problemàtica i el grau de satisfacció. Les biblioteques que ja havien dut

- a terme la digitalització dels seus catàlegs i col·leccions documentals eren la biblioteca de la Princeton University,⁶ i la Library of Virginia;⁷
- experiència de l'empresa en la gestió automatitzada de biblioteques i tractament de formats: es valorà especialment el coneixement del format CATMARC, ja que era essencial per a codificar correctament la informació de les fitxes. La mateixa empresa és el proveïdor del sistema informàtic utilitzat per la BC i disposa d'experiència provada en l'ús del format;
 - terminis d'execució del projecte: la solució proposada oferia resultats a curt termini, quant a la difusió de les col·leccions documentals de la BC i permetia assolir l'objectiu final, disposar dels registres en format CATMARC, a mig termini.

Una vegada feta la selecció calia disposar dels recursos econòmics necessaris. L'any 1996 la BC va presentar a l'Agència de Patrocini i Mecenatge de la Generalitat de Catalunya el projecte de reconversió perquè iniciés les gestions necessàries per a aconseguir una esponsorització. Com a resultat d'aquestes gestions la Biblioteca obtingué el patrocini de Telefónica de España S.A. que considerà d'importància rellevant i d'interès per a tota la comunitat científica, i en general per a tota la societat, la difusió del patrimoni bibliogràfic i documental i l'accessibilitat a la informació de les col·leccions documentals de la BC. Els termes i condicions d'aquest patrocini quedaren recollits en el conveni signat per ambdues institucions el mes de març de 1997.

4. DISSENY DEL PROJECTE

4.1. Identificació de l'objectiu i dels beneficis del projecte

L'objectiu del projecte Biblos és convertir en registres bibliogràfics en format estàndard llegible per màquina les descripcions bibliogràfiques de les fitxes dels catàlegs manuals de la BC incloent-hi les dades de localització de l'exemplar.

Els beneficis són clars i diversos: donar notícia de totes les col·leccions documentals de la biblioteca, més enllà de les seves parets, via Internet; convertir la informació bibliogràfica de les fitxes en registres codificats en format estàndard per a constituir una única base de dades en línia; facilitar l'accés i la

6. Princeton University, *Electronic Card Catalog* <<http://imagecat1.princeton.edu/ECC/>>.

7. The Library of Virginia, *Electronic Card Indexes* <<http://image.vtls.com/collections/>>.

còpia de registres a tota la comunitat bibliotecària; optimitzar els processos interns d'adquisició, de control de duplicats, de catalogació..., i agilitar els serveis d'obtenció de documents.

4.2. Àmbit del projecte

La segona decisió a prendre fou quins catàlegs calia escannear. Per a codificar la informació només era necessari disposar de les imatges de les fitxes principals dels catàlegs d'autors, ja que d'elles s'obtindrà la catalogació descriptiva i els punts d'accés però, per a facilitar-ne la consulta als usuaris i ja que els costos d'escanear eren relativament baixos en comparació amb els de codificació, es decidí d'escanear tant els catàlegs alfabètics d'autors com els alfabètics de matèries. D'aquesta manera, durant el període en el qual els catàlegs només seran accessibles a través de la base de dades d'imatges —és a dir, mentre es codifica la informació— els usuaris podran recuperar les fitxes a Internet buscant tant pels índexs d'autors com pels de matèries.

4.3. Definició de requeriments

Requeriments definits per la BC:

- les fitxes dels catàlegs de la BC no sortiran del recinte de la biblioteca per tal de no distorsionar la consulta als usuaris;
- la qualitat de les imatges es valorarà d'acord amb els criteris preestablerts conjuntament per l'empresa i la biblioteca; en cas que una imatge no compleixi els mínims de qualitat establerts caldrà procedir al reescanear de la fitxa sense cost afegit;
- l'empresa proporcionarà l'equipament específic per a l'escanear: escàners i els PC;
- l'empresa contractarà el personal necessari per a dur a terme l'escanear de les fitxes, en una primera fase i per a realitzar la codificació, en la segona;
- proporcionarà i instal·larà el programari de gestió d'imatges;
- l'empresa crearà, en segona fase, els registres en format CATMARC a partir de les plantilles de catalogació i de la informació de les imatges de les fitxes dels catàlegs manuals, corresponents tant a monografies com a publicacions en sèrie, d'acord amb les especificacions elaborades conjuntament;
- en crear els registres codificats informarà correctament el camp 856 per a enllaçar cada registre amb la imatge de la fitxa original.

Requeriments definits per VTLS:

- la BC proporcionarà l'espai i la infraestructura general necessària dins dels locals de la biblioteca per a la instal·lació dels equipaments i de les persones que han de realitzar l'escanueig (taules, cadires, telèfon, connexions elèctriques i informàtiques);
- la biblioteca disposarà d'un servidor, d'acord amb les característiques definides per l'empresa perquè hi resideixi la base de dades d'imatges;
- proporcionarà accés via telnet, ftp, http, per a disposar de suport remot des dels Estats Units;
- elaborarà els índexs per a cadascun dels catàlegs que s'escanuejaran, en format text;
- definirà les plantilles de catalogació en format MARC que reflectiran la distribució de les dades a les fitxes manuals;
- la BC designarà un interlocutor únic per a fer el seguiment global del projecte.

4.4. Calendari

La durada global del projecte es definí en quinze mesos, a partir de l'inici del procés d'escanueig. Els terminis aproximats previstos per a cadascuna de les fases foren:

Fase d'escanueig: Octubre 1997 - Gener 1998

Fase de codificació: Febrer 1998 - Gener 1999

5. REALITZACIÓ DEL PROJECTE

5.1. Accions preliminars

L'escanueig de les fitxes del catàleg de la BC presentava una problemàtica afegida: els *tracings* —a partir dels quals es generaran els punts d'accés dels registres MARC— es trobaven al revers de les fitxes i en llapis (difícilment reproduïbles per escàner). La BC contractà durant l'estiu de 1997 auxiliars que, conjuntament amb el personal del Servei d'Accés i Obtenció de Documents, va copiar els *tracings* a l'anvers de les fitxes; això subsanava dos problemes: d'una banda, el de la il·legibilitat d'alguns *tracings* —el llapis amb el temps s'havia anat esvaint— i, de l'altra, es podia escanuejar només l'anvers de les fitxes amb el consegüent estalvi de temps d'escanueig, d'espai en disc i de nombre d'imatges a recupe-

rar quan es cerca la informació. Aquesta ha estat una feina molt feixuga, però molt productiva.

D'acord amb els requeriments definits per l'empresa, abans de l'inici del procés d'escanueig, la BC va elaborar i lliurar a l'empresa els índexs en format text que, una vegada en html, havien de donar accés als conjunts de les imatges; adquirí un servidor UNIX amb prou capacitat perquè hi residís la base de dades d'imatges i definí les plantilles que havien d'ajudar a la codificació de la informació, de les quals parlarem més endavant.

5.2. Preparació dels catàlegs i adequació de l'espai i dels equipaments

El mes de setembre de 1997, tres especialistes de l'empresa van arribar a Barcelona per a realitzar la posta en marxa de Biblos. La Biblioteca cedí l'espai necessari de la seva seu per a instal·lar els equipaments i per a ubicar el personal que havia de realitzar l'escanueig. El fet d'escanuejar les fitxes dins del mateix edifici evità que els catàlegs sortissin de la Biblioteca i permetia que cada cedulari o fitxer del catàleg estigués menys de 48 hores no disponible.

La segona quinzena de setembre començaren a treballar els becaris contractats per l'empresa, tots ells estudiants de l'Escola Universitària de Bibliotecnologia i Documentació de Barcelona. Els becaris procediren a la revisió dels índexs per a detectar possibles errors ortogràfics o d'introducció de dades; a continuació, van preparar els catàlegs que s'havien d'escanuejar introduint senyalitzacions visuals per a cada conjunt de fitxes, i finalment assistiren a sessions d'aprenentatge en l'ús del software i del maquinari impartides per personal de l'empresa.

5.3. Fase d'escanueig

El dia 1 d'octubre de 1997, tal com s'havia previst, s'inicià la fase d'escanueig.

Característiques de les imatges: les imatges ocupen entre 5 i 15k cadascuna, són escanuejades en format TIFF comprimit, que el software de gestió d'imatges converteix en GIF quan es recuperen via Internet.

Característiques dels índexs: s'han establert quatre índexs en format html, un per a cadascun dels catàlegs alfabètics d'autors i alfabètics de matèries de la BC. Una vegada dins l'índex que es vol consultar, l'usuari podrà seleccionar la part de l'índex a la qual vol accedir (A, B, C...). Cada línia dels índexs indica el primer i el darrer encapçalament del conjunt de fitxes amb el qual enllaça. Cada conjunt conté al voltant de 100 imatges.

5.3.1. *Procediment de treball*

Treballaren simultàniament tres estacions d'escaneig amb tres torns diaris que cobrien des de les 8 del matí fins a les 11 de la nit. Les tasques que es desenvoluparen foren:

- preparació de les fitxes: obrir cedularis o fitxers i comprovar que les senyalitzacions del conjunt a escanear siguin les correctes. Preparar les fitxes que cal reescanear;
- escaneig de les fitxes: selecció del conjunt de l'índex html amb el qual s'associaran les imatges i escaneig;
- control de la qualitat de les imatges: visualització de les imatges per a acceptar-les, d'acord amb els criteris preestablerts, o rebutjar-les i marcar-les per a tornar-les a escanear. És en aquest punt que a les fitxes principals se'ls assigna una de les plantilles de catalogació predefinides.

El sistema generà de forma automàtica diversos informes perquè la Biblioteca pogués fer el seguiment del projecte: fitxes per reescanear, imatges rebutjades, progrés del projecte. Les imatges rebutjades definitivament corresponien normalment a fitxes molt deteriorades, i es rebutjaren perquè algunes de les dades eren il·legibles (per exemple el topogràfic).

Per tal de solucionar aquests casos, VTLS proporcionà a la BC un software que permet retocar les imatges manualment. A part del control de qualitat que realitza l'empresa, la Biblioteca inicià un control de qualitat directe sobre un, aproximadament, 10% de les imatges.

5.4. **Fase de codificació**

Una vegada finalitzada la fase d'escaneig, les imatges seran accessibles via Internet i començarà la fase de codificació en base a les imatges i amb l'ajuda de les plantilles predefinides.

Les plantilles s'han definit per a ajudar a reconèixer els diferents camps d'una fitxa (autor, títol, peu d'impremta...), en funció de la disposició de les dades, de l'ús de majúscules i minúscules, i d'altres signes tipogràfics utilitzats (claudàtors, parèntesis, etc.); per exemple, en una època determinada l'autor no es reproduïa a continuació del títol sinó que se substituïa per una ratlla, i després s'indicaven altres mencions de responsabilitat separades per punt espai, en una altra època se substituïa per punts suspensius...

S'han definit nou plantilles de catalogació: una per als documents descrits amb la puntuació ISBD, una altra per a les publicacions en sèrie i, set més per

a descriure formalment la resta de tipologies de fitxes. Amb l'ajut d'aquestes plantilles i a partir de les imatges, els operadors de l'empresa crearan manualment els registres en format MARC.

La catalogació descriptiva dels documents es codificarà en format CAT-MARC amb les dades corresponents als registres de fons per a les publicacions en sèrie. La codificació la realitzaran de forma manual operadors contractats per l'empresa. En un primer moment es valorà la possibilitat d'utilitzar tècniques d'OCR (reconeixement òptic de caràcters), però aquesta alternativa es desestimà, ja que el 45% de les fitxes són fetes a mà, amb grafies diferents, i podien generar una marge d'error molt elevat.

Es codificarà la informació bibliogràfica corresponent a camps de longitud variable i s'informaran només aquells camps fixos que es puguin extreure de la informació de la fitxa: llengua, data de publicació, país de publicació, tipus de document... Els registres bibliogràfics portaran associada la imatge escanejada de la fitxa a l'etiqueta 856.⁸ Periòdicament l'empresa lliurarà a la BC una cinta de registres perquè puguin ser revisats. Els registres convertits a format CAT-MARC es carregaran en una base de dades independent de la base de dades corrent, però igualment accessible via OPAC i Internet. Una vegada finalitzat el projecte Biblos, la BC preveu realitzar l'eliminació de duplicats, l'adequació dels punts d'accés i el traspass dels registres a la base de dades corrent, de forma gradual, en funció dels recursos disponibles.

6. CONCLUSIONS

El projecte Biblos es troba en curs de realització en el moment d'escriure aquest article (febrer 1998). Ha finalitzat el procés d'escaneig de les fitxes i d'aquí a pocs dies aquesta informació serà accessible a Internet. A partir d'aquest moment, la BC podrà iniciar una primera valoració dels beneficis que reporta la disponibilitat en línia dels catàlegs de la biblioteca en base a resultats quantificables, en relació amb els serveis als usuaris: increment del nombre de consultes via web, increment del nombre de peticions de préstec interbibliotecari; i en relació amb la millora de processos interns: control de duplicats, adquisicions, catalogació i obtenció de documents.

Una vegada finalitzada la conversió retrospectiva, la Biblioteca podrà oferir la còpia dels registres de la seva base de dades de fons retrospectius com a eina

8. Podeu trobar exemples de l'ús d'aquesta etiqueta a: *Guidelines for the use of field 856*. Rev. August 1997 <<http://lcweb.loc.gov/marc/856guide.html>>.

de conversió retrospectiva d'altres biblioteques, utilitzar-la per a l'intercanvi de registres amb altres institucions, etc. Tot i que la qualitat dels registres procedents de reconversió és sempre menor a la catalogació original amb el document davant, el fet que les dades es trobin en format estàndard llegible per màquina suposarà un valor afegit substancial als beneficis derivats de la difusió de les col·leccions.

El projecte Biblos ha permès posar en marxa un procés que la biblioteca amb el seu personal i amb els mitjans disponibles no hauria pogut realitzar en les condicions i terminis exposats. La utilització de serveis externs amb experiència en aquest camp, treballant en estreta col·laboració amb la BC i l'obtenció d'un patrocini global per al projecte ha permès dissenyar un projecte amb un inici i final clars, preveure un període de realització relativament curt i obtenir-ne uns resultats que creiem beneficiaran a tota la comunitat.