



12
JORNADES CATALANES
D'INFORMACIÓ
I DOCUMENTACIÓ
BARCELONA
19 I 20 DE MAIG DE 2010

Google Llibres i la Biblioteca de Catalunya

EUGÈNIA SERRA / Coordinació General. *Project Manager del projecte*
eserra@bnc.cat

SERGI FONT / Àrea de Direcció. *Day by day Manager del projecte*
sfont@bnc.cat

ELVIRA PERMANYER / Servei d'Emmagatzemament, Preservació i Conservació
epermanyer@bnc.cat

KARIBEL PÉREZ / Àrea de Tecnologia de la Informació
kperez@bnc.cat

RAMON NOVOA / Àrea de Tecnologia de la Informació
rnova@bnc.cat

Resum

La Biblioteca de Catalunya aposta clarament per fer accessible el coneixement des de qualsevol lloc, per a tothom i en qualsevol moment; gràcies a les noves tecnologies i en especial a la digitalització, aquest objectiu és cada vegada més factible.

El gener de 2007 la Biblioteca anunciava l'acord signat amb Google per a la digitalització dels seus fons en domini públic, i els d'altres quatre biblioteques patrimonials catalanes: Ateneu Barcelonès, Centre Excursionista de Catalunya, Biblioteca Pública Episcopal de Barcelona i Monestir de Montserrat.

En un temps amb molts reptes però no sempre massa recursos, la participació a *Google Llibres* ha representat i representa una oportunitat per donar a conèixer la pròpia institució, fer visible els fons de la biblioteca, difondre la cultura catalana, i facilitar el seu descobriment i coneixement a Internet per a qualsevol persona a través del portal de *Google Llibres*, sens dubte un referent quan es parla d'Internet.

Resumen

La Biblioteca de Catalunya apuesta claramente por dar acceso al conocimiento desde cualquier lugar, a todo el mundo y en cualquier momento; gracias a las nuevas tecnologías y en especial a la digitalización, este objetivo es cada vez más factible.

En enero de 2007 la Biblioteca anunciaba el acuerdo que había firmado con Google para la digitalización de sus fondos en dominio público, y los de otras cuatro



bibliotecas patrimoniales catalanas: Ateneu Barcelonès, Centre Excursionista de Catalunya, Biblioteca Pública Episcopal de Barcelona y Monestir de Montserrat.

En un tiempo de muchos retos pero no siempre demasiados recursos, la participación en *Google Libros* ha representado y representa una oportunidad para dar a conocer la institución, hacer visibles los fondos de la biblioteca, difundir la cultura catalana, y facilitar su descubrimiento y conocimiento en Internet por parte de cualquier persona mediante el portal *Google Libros*, sin duda un referente cuando se habla de Internet.

Abstract

The Biblioteca de Catalunya is clearly committed to giving everyone access to knowledge, anytime and anywhere; thanks to new technologies, and digitization in particular, this goal is more and more feasible.

In January 2007 the Biblioteca announced an agreement with Google to digitize its collection in public domain, and also those of four more Catalan heritage libraries: Ateneu Barcelonès, Centre Excursionista de Catalunya, Biblioteca Pública Episcopal de Barcelona and Monestir de Montserrat.

In a time of many challenges but not too many resources, to participate in Google Books is an opportunity of making known the institution itself, making visible the library funds, disseminating Catalan culture, and making easier their discovery and knowledge to anyone on the Internet through Google Books, certainly a reference when talking about the Internet.

Paraules clau

Digitalització; Biblioteques nacionals; *Google Llibres*; Biblioteca de Catalunya

Digitalización; Bibliotecas nacionales; *Google Libros*; Biblioteca de Cataluña

Digitization; National libraries; *Google Books*; Library of Catalonia

Introducció

El gener del 2007 la Biblioteca de Catalunya (BC) anunciava la signatura d'un conveni amb Google per a la digitalització i accessibilitat a milers de llibres de la seva col·lecció lliures de drets d'autor.

Aquest anunci era la culminació d'un procés iniciat mesos enrere en el que es van debatre, consensuar i fixar els termes del conveni de col·laboració.



Roda de premsa d'anunci del conveni (11 de gener de 2007)¹

La casualitat va voler que la història de la col·laboració entre la BC i Google comencés al maig del 2006, a unes *Jornades Catalanes d'Informació i Documentació* com les que celebrem aquests dies. Les jornades sempre han constituït un punt de trobada de professionals i agents del sector i en aquest cas no només va ser així sinó que va tenir com a resultat un acord que ha permès a la Biblioteca de Catalunya i a la cultura catalana fer un salt quantitatiu i qualitatiu en el seu posicionament al món virtual de la informació.

El 2006, feia poc més de 3 anys que Google havia posat en marxa el portal *Google Llibres*² (inicialment *Google Print*) bàsicament amb participació de biblioteques anglosaxones; per aquella època, estratègicament, Google havia girat els ulls cap a Europa amb la voluntat d'incrementar la seva oferta de continguts a altres cultures, llengües i documents.

¹ Nom i càrrec al 2007 dels presents a la fotografia: Josep Maria Martí (Director de la Biblioteca Pública Episcopal de Barcelona), Isabel Aguilera (Directora General de Google per a Espanya i Portugal), Joan Manuel Tresserras (Conseller de Cultura i Mitjans de Comunicació), Jens Redmer (Director del programa de Google Llibres a Europa), Dolors Lamarca (Directora de la Biblioteca de Catalunya), Oriol Bohigas (President de l'Ateneu Barcelonès), Manuel Castellet (Vicepresident del Centre Excursionista de Catalunya) i Damià Roure (Director de la Biblioteca del Monestir de Montserrat).

² *Google Llibres* <<http://books.google.es/bkshp?hl=ca&tab=wp>>



La Biblioteca de Catalunya per la seva part feia dos anys que havia iniciat accions intensives en l'àmbit de la digitalització selectiva. El *Pla Estratègic 2004-2008*³, vigent en aquell moment, definia entre les seves prioritats el context digital, tant pel que feia a serveis, com a la digitalització per preservar i difondre, com a la recollida de la web catalana, tot sota la visió de ser una biblioteca nacional "oberta, fiable i útil". D'acord amb aquesta estratègia havia impulsat diverses iniciatives com *ARCA (Arxiu de Revistes Catalanes Antiques)*⁴, *MDC (Memòria Digital de Catalunya)*⁵ i *PADICAT (Patrimoni Digital de Catalunya)*⁶.

La participació en *Google Llibres* encaixava exactament amb la visió de la BC i representava una oportunitat incomparable de digitalitzar de forma massiva el fons patrimonial antic de la Biblioteca, que d'altra manera, per una qüestió de costos i recursos, la Biblioteca difícilment hauria pogut assolir per ella mateixa. L'acord amb Google fa que el somni d'una biblioteca universal i gratuïta per a tothom sense esculls de cap mena estigui una mica més a prop.

1. Termes del conveni i planificació del projecte

1.1 El conveni

El conveni de col·laboració entre la BC i Google estableix la digitalització exclusivament de llibres en domini públic, és a dir, lliures de drets d'autors. En aquest punt Google és especialment curós i per tant es van acordar uns criteris força conservadors que veurem més endavant.

³ Biblioteca de Catalunya. *Pla Estratègic 2004-2008*. <http://www.bnc.cat/bc/qualitat/pestrategic2004_2008.doc>. A començaments del 2009 el Consell Rector de la Biblioteca de Catalunya ha aprovat el *Pla Estratègic 2009-2012* <http://www.bnc.cat/bc/qualitat/pestrategic_2009_2012.pdf> que continua amb les línies d'actuació definides a l'anterior pla.

⁴ *ARCA (Arxiu de Revistes Catalanes Antiques)* <<http://www.bnc.cat/digital/arca/index.html>> és un repositori digital cooperatiu obert i gratuït que actualment ofereix accés a més de 200 publicacions periòdiques que ja no es publiquen representatives de la societat i cultura catalana.

⁵ *MDC (Memòria Digital de Catalunya)* <<http://mdc.cbuc.cat>> és un repositori cooperatiu des del que es poden consultar, en accés obert, col·leccions digitalitzades de revistes catalanes antigues, fotografies, mapes, cartells, ex-libris, etc. relacionats amb Catalunya.

⁶ *PADICAT (Patrimoni Digital de Catalunya)* <<http://www.padicat.cat>> és un sistema d'informació que consisteix a capturar, processar i donar accés permanent a tota la producció cultural, científica i de caràcter general catalana produïda en format digital



Fruit de la digitalització s'obtenen dues còpies digitals, una de Google que és la que publica al seu portal *Google Llibres* i una altra de la Biblioteca, que la Biblioteca pot utilitzar en el seu propi lloc web o en d'altres amb els que coparticipi sempre que siguin iniciatives sense propòsit comercial. En el cas de la Biblioteca de Catalunya, es considera la còpia de Google com a còpia de difusió, i la còpia de la biblioteca es guarda a efectes de preservació.

El conveni defineix les responsabilitats de les parts, la Biblioteca selecciona i prepara els lots de documents d'acord amb els criteris establerts de comú acord i Google assumeix el transport dels documents fora de la Biblioteca, els costos de l'assegurança, la digitalització dels documents i la seva accessibilitat a través de *Google Llibres*.

Un aspecte important és que l'acord no estableix cap mena d'exclusivitat amb Google, per tant la Biblioteca pot tornar a digitalitzar per si mateixa o amb tercers qualsevol dels documents inclosos en el conveni si ho considera necessari.

El conveni inclou a més la digitalització, sota les mateixes condicions i criteris, dels documents patrimonials d'altres quatre biblioteques catalanes, fet que contribueix a incrementar el fons català a *Google Llibres* i a diversificar temàtiques i continguts, es tracta de les biblioteques de l'Ateneu Barcelonès, del Centre Excursionista de Catalunya, la Biblioteca Pública Episcopal de Barcelona i la del Monestir de Montserrat.

1.2 Planificació del projecte

Al llarg de prop d'un any des de l'anunci del conveni la Biblioteca i Google van dur a terme el procés de planificació que va consistir en la definició dels equips de treball, la quantificació i anàlisi dels documents a digitalitzar, l'establiment dels criteris d'estat de conservació i les característiques físiques dels documents, la identificació de circuits i processos, i la logística relacionada amb la preparació, transport i control dels lots de documents.

En aquest període es va realitzar la planificació i anàlisi ajustada no només del cas de la BC sinó també de les altres quatre biblioteques catalanes implicades en el projecte d'acord amb cada casuística particular.

El coneixement previ de Google fruit del seu treball amb altres biblioteques va ser fonamental per definir el projecte. Com a part de la planificació,



responsables del projecte de les biblioteques catalanes van tenir la oportunitat de visitar una de les biblioteques referents col·laboradora de *Google Llibres*, conèixer de primera mà la seva experiència i recollir les seves recomanacions i bones pràctiques.

2. Criteris i procés de selecció dels documents

El projecte *Google Llibres* a la Biblioteca de Catalunya s'articula en dues fases ben diferenciades. Una primera consistent en la selecció dels fons a digitalitzar i una segona que correspon a la digitalització dels fons seleccionats.

La primera fase comprèn les tasques prèvies a la digitalització pròpiament dita. Si bé l'aspecte central del projecte és la digitalització, totes les operacions que es duen a terme prèviament han de garantir que el resultat final sigui satisfactori. En aquest sentit la planificació de les tasques prèvies és fonamental per poder treballar amb rigor i a la vegada preveure les incidències sorgides durant el desenvolupament del projecte i possibilitar l'adopció de mesures correctores que facilitin l'assoliment dels objectius.

2.1 Selecció per dates a partir del catàleg

Donat que *Google Llibres* respon a un model de digitalització massiva calia establir un criteri segur i senzill d'aplicar que permetés fer una selecció ràpida i fiable a partir del catàleg de la Biblioteca.

Els documents a digitalitzar havien de complir certes característiques catalogràfiques. Havien de ser monografies impreses i havien d'estar lliures de drets d'autor. La legislació vigent en l'època dels documents establia (genèricament) que els documents serien en domini públic quan haguessin passat 70 anys de la mort de l'autor. Així es va acordar que els documents tinguessin data de publicació igual o menor a 1867, és a dir dues vegades el termini de drets d'autor per anar sobre segur -calculat al 2007 quan es va redactar l'acord.

Addicionalment, per incrementar el nombre de documents es van incloure els documents d'autors morts abans de 1927 i amb data de publicació anterior també a 1927, sempre que les dates de naixement i mort es trobessin informades al catàleg de la Biblioteca. Val a dir que la Biblioteca



de Catalunya, en tant que biblioteca nacional estableix unes fonts i procediments segurs abans de donar aquesta informació com a vàlida.

Al conjunt de documents resultants s'hi van aplicar després altres criteris relacionats amb la mida, l'estat de conservació, ... que van definir el conjunt de monografies a digitalitzar. Per localitzar els documents i aplicar-hi aquests criteris, a més de realitzar posteriorment tasques vinculades amb els lots de documents, era necessari disposar d'una base de dades i programari a mida que realitzés una sèrie de funcionalitats i oferís diferents subproductes. L'Àrea de Tecnologia de la Informació de la Biblioteca va desenvolupar una aplicació ajustada a les necessitats del projecte, les característiques de la qual veurem més endavant.

Així doncs amb els criteris de dates ja establerts es van extreure un seguit de dades bibliogràfiques mínimes i les dades relatives als exemplars per carregar-les a la base de dades feta a mida per a la gestió del projecte.

Donat que bona part dels documents d'aquest conjunt provenien del catàleg manual i havien estat creats a través de processos semiautomàtics de conversió, molts no disposaven de registre d'exemplar, ni per tant cap topogràfic associat. Per tant el primer pas va ser detectar per a quins d'aquest documents no existia registre d'exemplar.

Una vegada detectats, es va comprovar si alguns tenien etiquetes 949 (etiqueta local per descriure les dades d'exemplars en els processos de conversió del catàleg manual) per a crear automàticament l'exemplar a partir d'aquesta etiqueta.

Dels restants es varen cercar quins tenien etiquetes 993, etiqueta local que es feia servir a la BC per a incorporar la informació de l'exemplar en la pantalla de visualització del registre bibliogràfic (utilitzada amb VTLS, l'anterior sistema de gestió de la BC). I es van crear automàticament exemplars a partir d'aquesta etiqueta.

La resta de documents per als quals no es va trobar cap mètode per a crear un registre d'exemplar amb una data de topogràfic vàlida (uns 300), van ser descartats per a la digitalització a Google.

Així mateix es varen detectar topogràfics incorrectes "en blanc" o amb literals com "pendent per assignar" que es varen corregir o descartar.



Amb l'objectiu de facilitar les tasques de selecció dels documents, una vegada aconseguit el llistat d'exemplars a extreure, es van ordenar aquests per topogràfic i es va donar una seqüència nova de codis de barres a tots els documents. Cal recordar que la majoria dels documents de la Biblioteca es troben en dipòsits tancats i els topogràfics no corresponen a una classificació decimal, són topogràfics alfanumèrics que designen únicament la ubicació del document al dipòsit.

El format del fitxer d'extracció dels registres de la base de dades bibliogràfica va ser ISO2709 amb codificació de caràcters en Roman 8 amb format de dades bibliogràfiques CATMARC (en el moment de fer aquest selecció la biblioteca no havia fet encara el canvi de format a MARC21).

2.2 Selecció per característiques físiques i estat de conservació dels documents

Una vegada realitzada la selecció de documents lliures de drets d'autor, s'establiren altres criteris relacionats amb la mida i estat de conservació dels documents. Pel que fa a la mida es tractava de limitar a una determinada alçada i gruix del documents.

Quant a l'estat de conservació dels documents, Google s'ajusta als desitjos de cada biblioteca col·laboradora, que pot determinar ser més o menys restrictiva en la selecció en funció de l'estat físic dels llibres. Val a dir que el criteri de la Biblioteca de Catalunya ha estat d'intentar digitalitzar el màxim de documents per poc que el seu estat ho permetés, considerant que d'aquesta manera l'ús posterior de l'original es minimitza en bona mesura i s'eviten futures manipulacions que poden contribuir a malmetre el document.

Els criteris de selecció de la BC són bàsicament els següents:

- L'estat de les tapes i guardes del llibre han d'estar ben fermes al bloc del llibre, qualsevol deteriorament que impliqui despreniment de les tapes del bloc suposa el rebuig del document.
- L'estat del llom del llibre condiona la manipulació dels documents però sempre que el bloc del llibre sigui ferm és objecte de selecció malgrat que en ocasions sigui fràgil.
- L'obertura del llibre és un aspecte crític ja que malgrat l'estat de conservació del document si en obrir el document la pàgina impresa no



es pot veure sencera i queda tallada pel cosit de l'enquadernació implica el rebuig del document.

- La problemàtica del paper àcid és importantíssima en fer la selecció dels documents per ser digitalitzats i cal tenir-la molt present ja que la fragilitat del paper pot dificultar la manipulació del llibre i de les pàgines molt especialment en passar-les una a una. A efectes del projecte s'establiren una sèrie de comprovacions manuals de l'estat del paper del documents.

3. Metodologia i equip de treball

El desenvolupament de les tasques del projecte *Google Llibres*–Biblioteca de Catalunya ha estat possible gràcies a la implicació de diversos professionals de la Biblioteca de Catalunya especialistes en els àmbits diversos que engloba el projecte.

3.1 L'equip de direcció del projecte

D'acord amb Google calia establir un equip permanent per part de la Biblioteca amb perfils i responsabilitats concretes, així com diferents graus de dedicació quant a temps. Aquests perfils⁷ són:

- *Project manager*, és la responsable del projecte. És la interlocutora principal amb Google pel que fa al seu desenvolupament. D'acord amb les directrius de la directora de la Biblioteca de Catalunya ha dut a terme la planificació amb la col·laboració del *Day by Day Manager*, i ha donat forma a la col·laboració amb Google segons l'acord al que s'havia arribat.
- *Day by Day manager*. Coordina les tasques quotidianes pel que fa a l'organització i a la logística i resol les incidències que puguin sorgir durant la realització de les activitats relacionades amb el desenvolupament del projecte. Manté contacte constant amb Google i amb la resta de persones implicades en el projecte, en l'àmbit d'actuació que li és propi. Col·labora amb la *Project Manager* en la planificació.
- *SEPIC*, Servei d'Emmagatzemament Preservació i Conservació. El personal d'aquesta unitat de treball de la Biblioteca de Catalunya, encapçalat per la seva cap s'encarrega de la selecció dels documents a

⁷ S'han mantingut les denominacions de cada perfil en la seva forma anglesa ja que en la relació amb Google és la terminologia que s'utilitza.



digitalitzar en funció dels criteris establerts, quant al seu estat de conservació, i de la preparació dels lots de documents per ser digitalitzats.

- *ATI*, Àrea de Tecnologia de la Informació. Proporciona tot el suport tecnològic necessari per al desenvolupament del projecte en totes les seves fases: des de la selecció de documents del catàleg de la BC fins a la preparació dels lots, des del primer fins l'últim pas dels processos relacionats amb el projecte. Ha desenvolupat el programari de gestió necessari i altres millores tecnològiques vinculades al projecte.

A banda de l'equip permanent altres especialistes han contribuït al projecte, com ara la cap de la Unitat de Digitalització que ha assessorat la *Project Manager* a la presa de decisions quant a formats i estàndards de digitalització.

Per part de Google es definí també un equip de treball estable, del que cada membre actuava de forma especialitzada en un aspecte concret. Google establí una persona com a referent i responsable final del projecte.

3.2 Metodologia de treball

La metodologia per a preparar i executar el projecte va ser de general a concret. La definició i planificació va començar a través de diverses reunions presencials entre els equips complets de Google i la BC més representants de les altres biblioteques catalanes implicades per clarificar aspectes, criteris i objectius de caràcter general.

Posteriorment es va treballar cada aspecte del projecte de forma específica: criteris de selecció, aspectes tecnològics, logística i procediments. El canal de comunicació era habitualment el correu electrònic, si bé es mantenien reunions presencials puntuals, especialment quan es tractaren temes relacionats amb la logística, aspecte aquest que implicà la visita dels interlocutors de Google a cadascuna de les biblioteques en més d'una ocasió.

Una vegada iniciada la digitalització dels documents la *Project Manager* i el *Day by Day manager* han mantingut converses telefòniques periòdiques i simultànies amb diferents representants de Google per fer-ne el seguiment, primer cada quinze dies i posteriorment un cop al mes i/o quan ha estat necessari per tractar algun tema concret. A més a més hi ha hagut contacte permanent mitjançant el correu electrònic.



Per la seva banda la BC, en tant que interlocutora de les altres quatre biblioteques catalanes que participen al projecte, ha convocat periòdicament reunions amb als interlocutors d'aquestes biblioteques per prendre decisions conjuntes sobre criteris, per informar-los de l'evolució del projecte i ha estat en contacte més intensament en el moment que cadascuna de les biblioteques ha entrat en fase de selecció i/o digitalització.

3.3 Els equips de selecció

Tot el procés de selecció i senyalització dels documents en funció dels criteris consensuats entre Google i la BC sobre l'estat de conservació l'ha portat a terme un equip de bibliotecaris dedicats al cent per cent al projecte. Aquests equips s'han constituït gràcies a subvencions concedides pel Servei d'Ocupació de Catalunya (SOC).

Abans d'iniciar la selecció i senyalització els equips han rebut la formació específica necessària per part del Servei d'Emmagatzemament, Preservació i Conservació de la BC, que és qui també ha dut a terme el seguiment de la seva tasca.

En la sessió formativa es van donar els criteris per a la selecció quant a l'estat físic dels documents: la fragilitat del paper, la fermesa dels lloms o la dificultat de lectura degut a una obertura dels document difícil tant per la lectura com per a la digitalització, il·lustrant cada cas amb nombrosos exemples. Una vegada formats es van constituir equips de dues persones ja que es va considerar que era la millor manera de treballar amb llistes, llibres i codis de barres físics. L'objectiu era, doncs, fer el més àgil, eficaç i segur tot el procés de selecció.

El seguiment de la feina dels equips ha anat des de la coordinació d'aquests, la seva distribució en els dipòsits, passant per la localització física de les cotes topogràfiques dels documents en una primera fase, cal recordar que els dipòsits de la BC tenen 50 km. de prestatgeries, fins a resoldre problemes d'identificació d'autors o de títols segons les llistes de treball de què es disposava o la interpretació de cotes topogràfiques dels fons de reserva i les modificacions que s'haguessin de fer al catàleg de la BC.

4. Circuits i processos



4.1 Tasques de selecció

Els registres extrets del catàleg, com ja s'ha dit prèviament, es carregaren a la base de dades ad hoc desenvolupada per la BC que s'explica en detall al punt de programaris de gestió. El programari de la base de dades proporcionava unes llistes de treball a partir de les quals fer la selecció in situ amb el document al davant. Les llistes tenien els següents camps bàsics: topogràfics, autor, títol, codi de barres.

Les tasques del procés de selecció es realitzen document a document i consisteixen en:

- a. Comprovació de la mida dels documents.
- b. Avaluació de l'estat de conservació. Revisió física de tots i cadascun dels documents i aplicació dels criteris sobre l'estat de conservació consensuats entre Google i la BC exposats al punt anterior.
- c. Marcatge del document que complia tots i cadascun dels requisits mitjançant la inserció d'una cartolina verda en la qual s'hi anotava el topogràfic.
- d. Etiquetatge del codi de barres corresponent. Com s'ha explicat al punt de selecció de documents del catàleg per data, al conjunt de documents se'ls havia assignat al catàleg una seqüència nova de codis de barres, però calia enganxar el codi físicament a cada document. Val a dir que aquesta assignació havia de ser molt curiosa ja que el codi de barres és l'identificador utilitzat en tot el procés de selecció i digitalització.
- e. Marcatge dels documents que no complien els criteris establerts i eren rebutjats o que tenien alguna incidència que podia ser resolta posteriorment: llibres *in tonso*, topogràfic erroni, autor/títol no coincident amb la llista; en aquests casos els documents eren senyalitzats físicament amb una cartolina vermella i les incidències marcades a les llistes de treball a les columnes previstes en les quals es podia indicar cada casuística: no trobat, fora de mides, malmès i observacions.

Un cop fets tots aquests passos els documents estan a punt per ser digitalitzats. Quant als documents que tenen incidències, no s'han localitzat o que s'ha detectat algun problema d'identificació es fan els passos necessaris per resoldre el problema i esperar la seva selecció en un segon torn.

4.2 Digitalització dels fons



Un cop enllestides les tasques prèvies s'aborda la segona fase del procés que, de fet, constitueix el seu nucli. Abans de la digitalització estricta dels documents cal dur a terme un conjunt d'accions, que es detallen a continuació:

- a. El primer pas d'aquesta segona fase consisteix en recollir dels prestatges del dipòsit de la Biblioteca els llibres senyalitzats amb la cartolina verda que indica que són aptes per ser digitalitzats en el marc del projecte *Google Llibres*- Biblioteca de Catalunya. Els llibres són retirats del seu lloc deixant com a testimoni la cartolina verda que porta anotat el topogràfic.
- b. Als documents retirats se'ls llegeix el seu codi de barres (el seu identificador únic) i s'assegura que consten a la llista de documents. Els documents són aplegats en lots petits i dipositats en carros dissenyats específicament per a aquest propòsit. La lectura del codi de barres de cadascun dels documents s'introdueix a l'aplicació que gestionarà els lots per digitalitzar. En aquest moment, a més, es contrasta la informació de què disposa l'aplicació de gestió dels lots (títol i autor) sobre els documents amb el document físic, de manera que es poden detectar, si fos el cas, errors d'identificació. Així doncs, el codi de barres permetrà identificar el document i saber en tot moment el punt del procés en el qual es troba. Els lots es componen, doncs, de codis de barres agrupats en lots més petits, o el que és el mateix, d'un nombre indeterminat de documents que pot variar en funció del seu volum, col·locats en carros.
- c. En finalitzar aquesta part del procés el lot complet està en disposició de ser lliurat a Google per ser digitalitzat.
- d. Un cop Google finalitza les tasques de digitalització retorna el lot a la BC i els llibres poden ser col·locats de nou a la prestatgeria.
- e. Finalment, el llibre pot ser consultat a *Google Llibres*

5. Programaris de gestió i aspectes tecnològics

5.1 Programari de la BC

L'Àrea de Tecnologia de la Informació va desenvolupar una aplicació a mida per donar suport als processos implicats en el projecte. Els objectius de l'aplicació eren bàsicament tres:

- Fer un control dels documents que s'han d'enviar a digitalitzar, per tal de



poder etiquetar-los com a disponibles o descartar-los en cas de no complir els requeriments.

- Permetre crear lots amb la documentació requerida per Google.
- Permetre a Google accedir a la informació bibliogràfica i dels lots per tal de que la puguin introduir a la seva aplicació.

5.1.1 Components de l'aplicació

L'aplicació desenvolupada a la BC consta de dos components. Un component és en línia amb les dades dels documents seleccionats a partir del catàleg per digitalitzar i està dissenyat per a fer servir des de l'escriptori. L'altre és fora de línia i permet anar pels dipòsits seleccionant físicament els documents. Cada component depèn de l'altre: per a poder fer servir el programa de dipòsits necessitem dades que generem amb la part en línia, i aquest alhora en genera informació que necessitem per a preparar els lots.

A. Component en línia

És web, desenvolupat en PHP sobre Linux, Apache i MySQL (LAMP) i treballa en línia amb les dades dels documents. Aquesta part alhora es subdivideix en tres grans blocs:

- 1. Preparació del documents: permet marcar els documents que es poden digitalitzar.**
 - Càrrega de les dades del catàleg a l'aplicació. A partir del fitxer en format ISO2709 es va fer una conversió bàsica a format MARC21 (en el moment de càrrega hi havia diverses institucions del projecte, entre elles la BC, que feien servir format CATMARC) i es generen els registres d'exemplar.
 - Generació de llistats. Es creen una sèries de llistats en format PDF per poder ser impresos amb la informació necessària per a que el personal que ha de revisar l'estat dels documents pugui anar als dipòsits i anotar si tot està correcte o hi ha alguna anomalia. Els criteris de generació són que no hi hagin més de 250 exemplars per llistat i que estiguin, en la mesura del possible, ordenats segons es trobaran ubicats als dipòsits.
 - Actualització de dades. Permet recuperar i veure la informació en línia dels exemplars de cada llistat i actualitzar la informació segons el que s'ha anat anotant al dipòsit.



II. Generació de lots: permet gestionar els documents que s'enviaran.

- Generació dels lots per a una data concreta, indicant el nombre de carros que els componen.
- Compleció dels lots amb els codis dels documents que es posen físicament als carros.
- Generació dels fitxers necessaris per a que Google pugui tractar els lots i un fitxer per a poder enviar a l'entitat asseguradora els documents amb el contingut de cada lot.

III. Accés a Google: accés remot a Google de les dades dels lots

Un cop generada la documentació requerida per Google es crea un punt d'accés de forma que Google pugui accedir remotament, tant per consultes puntuals com per processos automàtics de recollida en lots d'informació.

B. Component fora de línia

Consta d'una petita aplicació client, desenvolupada en PHP i GTK, que es fa servir a l'hora d'omplir els carros de cada lot. Aquesta aplicació ha de presentar les següents característiques:

- Treballar fora de línia ja que s'ha de fer servir als dipòsits on no sempre hi ha accés a la xarxa. Malgrat això permet carregar un fitxer de text que es poden obtenir a partir del primer component amb les dades bàsiques de comprovació (títol, autor).
- Ser fàcil d'utilitzar. El programa carrega la informació bàsica del lot i es pot fer servir amb un conjunt molt reduït de botons i un lector de codi de barres.
- Generar un fitxer amb la informació dels documents que s'inclouran al lot corresponent.

Quan es va dissenyar l'aplicació es van tenir en compte una sèrie de requeriments que després han resultat no adaptar-se del tot a alguns aspectes de la manipulació final dels documents. Malgrat això l'aplicació i el disseny de les dades han resultat suficientment flexibles per a afegir petites modificacions per adequar el funcionament a les necessitats reals, i a la



vegada realitzar petits desenvolupaments per tal de poder explotar les dades d'una manera que no s'havia previst inicialment.

5.2 L'aplicació de seguiment de Google

Google ha desenvolupat una aplicació per al seguiment del procés de digitalització. Aquesta aplicació permet, a les biblioteques que participen en el projecte *Google Llibres*, el següent:

- Dóna accés a informació detallada sobre l'estat del procés de digitalització dels documents. Es pot fer una consulta individual, per codi de barres, o bé avançada, filtrant la cerca en funció de diferents paràmetres. El resultat obtingut detalla si el document ha estat digitalitzat i altres dades referents al procés.
- Permet descarregar les còpies en alta resolució dels documents digitalitzats que resten a la BC.
- Proporciona informes globals sobre els processos, agrupats cronològicament, que es refereixen a l'estat de digitalització dels documents per fer-ne el seguiment.

Aprofitant les característiques de l'aplicació de Google, l'Àrea de Tecnologia de la Informació (ATI) ha desenvolupat un procés per lots que permet programar la descàrrega de manera automatitzada de les còpies d'alta resolució.

5.3 Altres millores tecnològiques vinculades al projecte

A partir de les aplicacions i funcionalitats que ofereix Google, l'ATI amb les eines adients ha desenvolupat millores tant per a la web com per al catàleg en línia de la BC.

A la web de la BC ha incorporat una pàgina que permet cercar únicament als documents de la BC dins del projecte *Google Llibres*.



Pàgina d'accés als llibres de la BC digitalitzats a Google Llibres

A partir de l'aplicació de Google, l'ATI obté els llistats de codis de barres per als quals la digitalització ja es troba disponible a *Google Llibres*, amb aquesta informació actualitza el catàleg en línia de la BC, enllaçant-hi a cada registre la seva digitalització i incorporant-hi la portada al costat del registres, de manera que s'enriqueix el catàleg i es proporciona a l'usuari una forma senzilla d'accedir al document digitalitzat a text complet.

Pàgina de resultats: Anterior Següent

Autor corporatiu Catalunya. Corts (1599 : Barcelona)

Títol Capitols del General del Principat de Catalunya, comptats de Rossello y Cerdanya fets en les Corts celebrades en lo monestir de Sant Francesch de Barcelona per la S.C.R.M. del Rey don Phelip ... per lo redres del General y Casa de la Deputacio en lo any M. D. XCIX

Publicació Estampats en Barcelona : de manament dels senyors deputats del General de Catalunya : en la estampa de Jaume Cendrat, 1601

Descripció [24], 192, [8] p. ; 4º ; 21 cm

Exemplars **Registre Complet** [Enllaç permanent a aquest títol](#)

Localització	Topogràfic	Volum	Còpia	Consulta de	Tipus préstec	Nota
Dipòsit de Reserva	Mar. 125-8º			SALA RESERVA	NO ES PRESTA	Enq. perg., taques d'òxid i d... (+)

Exemple de vinculació entre el catàleg de la BC i el document a Google Llibres



6. Primers resultats

A primer trimestre de 2010 és previst que s'hagi digitalitzat aproximadament un 60% del total de documents seleccionats a les cinc biblioteques involucrades al projecte. En el cas de la Biblioteca de Catalunya el percentatge és de gairebé el 80%.

Més de 50.000 documents de la BC estan disponibles per a la seva consulta o descàrrega a *Google Llibres*. La Biblioteca de Catalunya, l'Ateneu Barcelonès i el Centre Excursionista de Catalunya tenen digitalitzats o en curs de digitalització la pràctica totalitat dels seus fons seleccionats, i la Biblioteca Pública Episcopal i la del Monestir de Montserrat es troben a l'inici del procés, si bé en breu disposaran de documents ja accessibles.

Tot i que en el moment de presentar aquesta comunicació (gener 2010) no disposem d'estadístiques periòdiques d'ús, els mostrejos realitzats sobre els documents digitalitzats a *Google Llibres* de la Biblioteca de Catalunya proporcionen xifres molt satisfactòries amb un nivell de consulta alt, a tall d'exemple en una sola setmana del mes d'octubre passat s'havien consultat un 28% dels documents de la BC.

7. Conclusions

Per a la BC col·laborar amb Google ha estat una experiència totalment positiva i enriquidora. *Google Llibres* és una eina de descobriment que, com bona part dels productes de Google, té un fort impacte social i cultural i és d'ús corrent per part de sectors d'usuaris habituats a un ús intensiu d'Internet però difícils de captar per a la Biblioteca.

Des del punt de vista metodològic i de planificació, treballar amb l'equip de Google ha proporcionat a la Biblioteca una experiència molt interessant; les institucions culturals estan en general en poca relació amb el sector privat i conèixer els seus procediments, hàbits i protocols proporciona idees i visions diferents que poden ajudar a millorar el funcionament intern de la Biblioteca.

Cal destacar el resultat més visible del projecte que és l'increment de l'oferta d'informació i l'accessibilitat a les col·leccions de la Biblioteca dut a terme en un període de temps relativament curt i amb una inversió molt petita pel que fa a pressupost. Atenent a la disponibilitat de recursos i pressupost de les



biblioteques al nostre país, la relació de l'estalvi -que equival al pressupost sencer de la BC en un any- amb els resultats i la seva qualitat, no és comparable amb cap altra iniciativa que hagi emprès la Biblioteca abans, i ha de constituir un precedent per accions futures de naturalesa similar. La col·laboració entre el sector públic i el privat ha de ser en els propers anys necessària i desitjable en un context de beneficis que satisfaci a les parts implicades.

Finalment, constatar que les administracions i el sector públic de la cultura han de promoure models d'inversió col·laboratius sostenibles sense oblidar l'objectiu i l'essència dels serveis públics, que són i existeixen per a proporcionar a les persones l'accés a la informació i al coneixement en un marc de qualitat i gratuïtat.

Bibliografia

Biblioteca de Catalunya. *Biblioteca de Catalunya* [en línia]. Barcelona: BC, 2005- . <<http://www.bnc.cat>> [Consulta: 24 març 2010].

Google Llibres [en línia]. Beta [versió]. [Mountain View, Calif.]: Google. <<http://books.google.es/bkshp?hl=ca&tab=wp>> [Consulta: 24 març 2010].

Proyecto para bibliotecas de la búsqueda de libros de Google [en línia]. [Mountain View, Calif.]: Google. <<http://books.google.es/intl/es/googlebooks/library.html>> [Consulta: 24 març 2010].

"*Projecte Google*" [en línia]. *BCNews*. Gener 2007. <<http://www.bnc.cat/noticies/aquestmes.php?data=012007>> [Consulta: 24 març 2010].