

Digitalització dels llibres de les biblioteques

Miquel TÉRMENS i GRAELLS

RESUM: *Es descriuen els diversos projectes d'abast global que darrerament s'han presentat per digitalitzar de forma massiva els llibres de les biblioteques. S'analitzen els objectius d'aquest tipus de projectes i els aspectes tecnològics i d'organització que comporten. Finalment, es comenten les implicacions que aquests nous serveis poden tenir pels que ja proporcionen les biblioteques i en la funció social d'aquests centres.*

MOTS CLAU: projectes de digitalització; digitalització; llibres; biblioteques digitals; Google Print Library; Open Content Alliance; Open Library; Universal Library; Biblioteca Digital Europea; Project Gutenberg; World Digital Library; Amazon; European Library; British Library.

E-BOOKS, LLIBRES ELECTRÒNICS I LLIBRES DIGITALITZATS

Les bases de dades i les revistes són realitats consolidades en el medi digital, però aquest no és el cas dels llibres. La seva translació digital es troba fragmentada i s'utilitzen diversos termes, no sempre de manera correcta, per referir-s'hi. Veurem a continuació que disposem de tres termes relacionats però que identifiquen conceptes diferents.

Els llibres electrònics és refereixen al contingut; es tracta de presentar el contingut intel·lectual, però a vegades també artístic i estètic, dels llibres impresos tradicionals en forma digital. Aquests continguts es poden consultar de diverses maneres: mitjançant aparells electrònics específics –els *e-books*–, que d'alguna manera recorden les formes dels llibres impresos, o mitjançant dispositius informàtics com els ordinadors i les agendes electròniques –PDA, etc. Quant a la forma de distribució, els llibres electrònics es poden obtenir de manera tangible, gravats en suports òptics com el CD-ROM o el DVD, o a través de la xarxa Internet, sigui descarregant-los d'algun servidor o usant-los de manera interactiva.

Miquel Térmens i Graells. Universitat de Barcelona. Departament de Biblioteconomia i Documentació. termens@ub.edu

Article arribat el 8 de febrer de 2006

El concepte de llibre digitalitzat es refereix al procés de creació d'un llibre electrònic que s'ha generat a partir de l'escaneig i el processament posterior d'un llibre imprès preexistent. Parlant de manera precisa, un llibre digitalitzat és antagònic a un llibre creat digitalment des de l'inici, i ambdós són llibres electrònics.

Donats aquests tres conceptes –*e-books*, llibres electrònics i llibres digitalitzats–, des d'inicis de la dècada del noranta, i a vegades fins i tot abans, se n'han fet desenvolupaments i se n'ha debatut de manera continuada les respectives viabilitats i repercussions. Pel que fa als *e-books* el problema ha estat tecnològic i industrial: és possible crear un aparell que simuli les formes dels llibres de paper (mida, pes, ergonomia, etc.) i n'augmenti les prestacions (cercabilitat, capacitat, navegació, etc.) gràcies a l'electrònica?; es podrà vendre als usuaris?; els lectors ho acceptaran?; hi haurà mercat? Pel que fa als llibres electrònics el problema només ha estat de prestacions i de viabilitat comercial. Primer, el debat, als anys noranta, es va centrar en refinar les tècniques de digitalització, després en la creació dels sistemes d'accés (portals bibliotecaris, per exemple) i més tard en les polítiques de digitalització, enteses com la prioritització dels fons a digitalitzar depenent del finançament disponible.

No és l'objectiu d'aquest article detallar les vicissituds de les tres vies al llarg dels darrers anys, però sí que és bo saber que els *e-books* encara no han resolt els seus reptes tècnics i només recentment el sistema Librie de Sony ha trobat la viabilitat comercial en el mercat japonès. Els llibres electrònics han tingut dificultats per vertebrar el seu mercat, definint unes línies determinades de producte i arribant a uns usuaris tipus; per ara, s'ha comercialitzat principalment el llibre tècnic, l'acadèmic i el de referència com ara manuals universitaris, diccionaris i enciclopèdies. A finals de la dècada dels noranta es va començar a desplegar la comercialització, mirant de vendre a les biblioteques universitàries paquets de llibres electrònics de manera similar a com es feia amb les revistes electròniques. Netlibrary es va revelar com l'empresa líder, però la crisi dels negocis *puntcom* va fer que fes fallida i que fos comprada per OCLC. Pel que fa als llibres digitalitzats, es tracta d'iniciatives que pel seu cost i per la necessitat de disposar dels fons originals han quedat dins de l'àmbit dels projectes públics; cal assenyalar, però, que hi ha hagut petites experiències de gestió comercial, com la pionera de l'empresa californiana Octavo,¹ que actualment encara ven en CD-ROM una col·lecció de llibres clàssics digitalitzats.

1. <<http://www.octavo.com/>> [Consulta: 7-2-2006].

DIGITALITZACIÓ: PER QUÈ ARA?

En els darrers temps s'ha posat de moda parlar de la digitalització dels llibres de les biblioteques. Per a la gent de la professió aquesta idea no és nova, però sí que ho és que s'hagi convertit en un tema habitual dels mitjans de comunicació de masses. És sabut que aquesta popularitat es va iniciar a finals de 2004 amb l'anunci de Google d'emprendre un gran projecte en aquesta línia, però això no ho explica tot. Aquest article vol sistematitzar les informacions que tenim disponibles sobre la qüestió i intentar mostrar què hi ha més enllà del que ens diuen els diaris i els programes de televisió.

Primer de tot hem de tenir clar que la iniciativa de Google és important, però més com a desencadenant del fenomen actual que no pas com a actuació realment innovadora. La digitalització massiva dels llibres de les biblioteques es planteja ara perquè ara és plantejable, perquè mai abans havia estat viable, això és l'important. Analitzem breument a continuació els canvis que s'han donat.

Des de fa anys l'economia mundial ha entrat en una nova fase: s'ha creat un mercat mundial, tant de productors com de consumidors; en aquest canvi hi han tingut un paper fonamental els augments continuats de la productivitat, l'aplicació intensiva de les noves tecnologies de la informació i la comunicació i, finalment, la concentració de capitals financers.

Aquest canvi general també està tenint importants efectes en el sector de la gestió de la informació, no en va hom parla del fet que s'està creant la societat de la informació. Un d'aquests efectes és la facilitat per explotar el patrimoni documental gràcies a les eines informàtiques. Fins fa ben poc, la majoria d'organitzacions consideraven l'arxiu com una càrrega, una font de despeses que calia suportar bàsicament per raons legals (fiscals, registre de drets) i històriques. La informàtica permet ara fer una explotació intensiva dels continguts de l'arxiu i convertir-lo en una nova font de negoci corporatiu. Uns dels primers que ho ha entès és el món editorial, amb la generació de nous productes com la consulta de l'hemeroteca dels diaris i les edicions barates de llibres descatalogats.

Aquest interès empresarial ha estat possible gràcies al desenvolupament de la informàtica. El més important pel que fa al nostre tema d'estudi ha estat la capacitat d'emmagatzematge: la digitalització de grans volums de premsa, de llibres, de fotografies o d'altres materials documentals requereix disposar d'unes capacitats d'emmagatzematge electrònic mai vistes. L'emmagatzematge no sols és un problema de maquinari, sinó també d'organització lògica de la informació, de disposar de les eines per a la recuperació de la informa-

ció i de seguretat de tot el sistema (fiabilitat davant de contingències, seguretat contra intrusos, servei 24x7). Segons l'Internet Archive, l'escaneig d'un milió de llibres ocupa 6 petabytes,² volum que es pot comparar amb els 1,7 a 5 petabytes que hom calcula que ocupen les bases de dades Google de buidatge del web. Els sistemes més avançats ja permeten muntar un sistema d'emmagatzematge d'1 petabyte per «només» 2 milions de dòlars.³

Aquesta immensa capacitat d'emmagatzematge no serveix de res si no es pot processar; en aquest cas no es tracta de disposar d'una alta capacitat de computació, com en els sistemes de supercomputació orientats a la recerca, sinó de poder gestionar un gran nombre d'objectes digitals i de transaccions d'usuaris. La programació orientada a objectes i la utilització de l'XML com a llenguatge estàndard d'intercanvi i de manipulació de dades han estat fonamentals des de finals de la dècada dels noranta per crear grans bases de dades i per oferir serveis des de sistemes informàtics diferents. Pel que fa a objectes digitals l'avanç potser és encara major, ja que les tècniques de digitalització retrospectiva i d'OCR que es van començar a assajar de manera sistemàtica a partir de 1994 ara ja estan consolidades i es disposa d'un catàleg coherent de bones pràctiques a aplicar.

Després de tots aquests desenvolupaments encara quedaven dos problemes per resoldre: la digitalització eficaç d'obres enquadernades i la preservació dels objectes digitals. Actualment, la digitalització de fulls solts es pot fer de manera automàtica, amb escàners que incorporen alimentadors de paper, però quan es tracta de llibres fins ara es tenien dues opcions: o digitalitzar-los manualment, passant les pàgines a mà, o bé, si és possible, guillotinant el volum per fer-lo passar per l'alimentador automàtic; no cal dir que si la primera opció és molt costosa, la segona és un atemptat contra la integritat del llibre; davant d'aquest panorama, la primera opció se sol deixar per a obres úniques o antigues i la segona per a obres més modernes. Darrerament, però, s'han fet desenvolupaments de cara a construir màquines per digitalitzar automàticament llibres sense desenquadrar-los, dotant-les amb un delicat mecanisme que els permet passar les pàgines d'una en una. En aquests moments es comercialitzen dues màquines: la suïssa 4DigitalBooks⁴ i la nord-americana Kirtas;⁵ també se sap que Google n'ha creat per al seu ús. No cal dir que es tracta de màquines molt cares, però que gràcies a la seva elevada

2. 1 petabyte = 1.000 terabytes = 1.000.000 gigabytes = 1.000.000.000 megabytes. Més informació a: <<http://en.wikipedia.org/wiki/Petabyte>> [Consulta: 7-2-2006].

3. <http://news.zdnet.com/2100-9584_22-5808754.html?tag=nl> [Consulta: 7-2-2006].

4. <<http://www.4digitalbooks.com/>> [Consulta: 7-2-2006].

5. <<http://www.kirtas-tech.com/>> [Consulta: 7-2-2006].

productivitat (fins a 3.000 pàgines per hora) compensen la inversió i, el que és més important, ningú no es plantejaria projectes de digitalització de centenars de milers de llibres si no existissin.

El segon problema pendent parla del futur: quin sentit té invertir grans quantitats de diners i crear nous serveis si no podem assegurar la conservació d'aquests nous objectes digitals? La preservació digital és un tema encara no resolt però que en els darrers dos anys, almenys pel que fa a les tècniques, s'hi comença a veure la llum; en altres paraules, encara no es disposa de tècniques de preservació assentades i provades, però sí que es disposa de prototipus i de sistemes pilot prou avançats com per preveure l'obtenció de solucions en pocs anys. Davant d'aquest panorama no és estrany que moltes institucions públiques i empreses privades enduguin grans projectes de digitalització, atès que «només» han de preveure dissenyar sistemes prou oberts per migrar a mig termini cap a les solucions que s'estableixin i aquest és un requeriment que s'assoleix amb diners i amb una bona direcció.

A banda dels requisits tecnològics, tres altres condicions han estat necessàries. La primera i més òbvia ha estat l'existència d'empreses amb una alta capacitat tecnològica i financera disposades a emprendre projectes de llarga durada. Tenen unes característiques ben peculiars: empreses amb molts diners, disposades a invertir-los en projectes nous, de tipus cultural (aparentment no lucratiu), nous en l'aspecte tècnic i amb resultats a llarg termini. Són projectes de risc, però també de prestigi i, encara més, de caire estratègic si es vol dominar una part important dels continguts que en el futur circularan per les xarxes informàtiques. Ara aquestes empreses existeixen i es diuen Google, Microsoft, Yahoo, Amazon, Hewlett Packard, Internet Archive, etc.

La segona condició no tecnològica ha estat de tipus organitzatiu: no parlem de projectes que es facin amb un amb un equip de cinc persones i un parell d'escàners i d'ordinadors personals; parlem de projectes que s'executen simultàniament en diverses ciutats o, fins i tot, a països diversos, amb centenars de persones treballant-hi, amb furgonetes dedicades a portar cada dia uns milers de llibres des de la biblioteca fins a la nau industrial on es desenvolupa el procés. Potser el servei resultant serà de tipus cultural, però el procés ha de ser totalment industrial, ha calgut la maduració de processos de *work-flow*, de control de la productivitat i de control de la qualitat, adaptats del món industrial.

La tercera condició ha estat de mentalitat. Ha calgut un canvi de mentalitat a les institucions dipositàries de documentació, que han transformat la seva antiga visió exclusivament patrimonialista i conservadora en una altra que es focalitza en les activitats de serveis, siguin portades a terme de manera di-

recta o bé mitjançant aliances amb agents externs. Això s'ha pogut fer després del reconeixement de les pròpies debilitats però també de les fortaleces: debilitats perquè ha calgut assumir que sense ajudes externes les institucions culturals mai no podrien digitalitzar el seu patrimoni a gran escala, més enllà d'alguns documents i col·leccions triades,⁶ però també la consciència que aquest patrimoni només el tenen elles.

COM ÉS UN PROJECTE DE DIGITALITZACIÓ DE LLIBRES

La digitalització de llibres en si mateixa només és una tècnica o, si es vol, un conjunt de tècniques, però la seva caracterització definitiva ve sobretot marcada pels objectius que s'hagin definit i, en part, també pel volum de la documentació que s'ha de tractar. Anem a resseguir tot seguit les principals fases del projecte.

- a) *Objectius de servei*. El servei que es construeixi pot anar orientat a públics diversos: en el cas d'una audiència especialitzada (per exemple, orientada a la recerca) caldrà una digitalització d'alta qualitat, encara que això s'aconsegueixi a costa d'una reducció de la velocitat d'accés; en canvi, si es va cap a una audiència general (per facilitar la lectura o el coneixement de materials fins llavors no tan accessibles), serà important basar-se en formats estàndard que es puguin llegir des de qualsevol ordinador i disposar també d'un sistema ràpid. També cal decidir si els llibres s'ofereixen en forma d'imatges o bé de text cercable: si és així, caldrà aplicar processos d'OCR; una bona mesura del nivell de qualitat d'un projecte de digitalització és el grau d'èxit del reconeixement òptic de caràcters, saber si les errades són esmenades a mà i com el text resultant és marcat amb etiquetes. També cal preveure si en el futur es voldran generar subproductes com ara reproduccions facsímil.
- b) *Estàndards*. És imprescindible treballar sobre estàndards, pel que fa a formats dels fitxers, a llenguatge d'etiquetat i a interoperabilitat. Les pàgines dels llibres, un cop escanejades, se solen guardar en format TIFF no comprimit i d'alta resolució, a partir del qual es poden generar, *offline* o de manera interactiva, les imatges de resolució mitjana i baixa en for-

6. La Universitat de Michigan ha reconegut que només tenia capacitat per digitalitzar 5.000 llibres l'any mentre que l'aliança amb Google els pot portar a digitalitzar més d'un milió de llibres l'any; calculen que s'han estalviat d'estar escanejant durant 1.600 anys. <<http://www.lib.umich.edu/staff/google/public/faq.pdf>> [Consulta: 7-2-2006].

mat JPEG o PNG que es transmeten als usuaris. Normalment, els fitxers obtinguts inicialment són tractats de manera automàtica amb algun programa de retoc gràfic a fi de millorar i homogeneïtzar característiques visuals com ara la gamma de colors i el contrast. Quant a programació, l'XML s'està imposant gràcies a la seva flexibilitat. També cal elegir algun estàndard de metadades (METS, MODS, Dublin Core, etc.) i, si cal, de marcatge del text (TEI o altres).

- c) *Processos. Workflow.* En un projecte es tracten milions de fitxers, molts dels quals estan interrelacionats entre ells (per exemple, els corresponents a les pàgines d'un mateix llibre, o els de les diverses resolucions gràfiques d'una mateixa pàgina); també hi intervenen multitud de professionals, sovint a diverses localitzacions físiques. Els fitxers i els programes informàtics estan distribuïts en diversos servidors i poden canviar d'ubicació segons ho demani el procés de treball. També cal controlar els mateixos llibres físics que es van escanejar per evitar que es malmetin o es perdin, però també per assegurar que tornin en el menor temps possible a les sales de lectura o als dipòsits de procedència. Tots aquests són alguns dels condicionants que fan que sigui imprescindible dissenyar uns bons processos de treball i dotar-se d'eines informàtiques que permetin el control d'activitats, d'accions i d'objectes digitals i que en facilitin la traçabilitat.
- d) *Control de qualitat.* Un dels objectius del control de processos ha de ser el control de la qualitat: amb un volum de feina de les proporcions que s'ha descrit, un mal disseny tècnic o una mala execució automatitzada poden donar com a resultat desenes de milers de fitxers inservibles per allò que van ser generats. Cal que totes les fases del treball comptin amb controls específics de qualitat; alguns dels punts que cal vigilar són: no malmetre els llibres, no deixar-se pàgines sense escanejar, no perdre la seqüència dels fitxers, generar metadades correctes, lligar les metadades amb els fitxers corresponents, calibrar l'OCR al màxim nivell possible, minimitzar les actuacions manuals, controlar el *checksum* dels fitxers, etc.

En els projectes de digitalització de llibres també s'ha de vigilar que el personal implicat tingui la formació necessària i s'ha de mantenir un control estadístic del desenvolupament general de les activitats. Els problemes de tipus no tècnic també són fonamentals, entre els quals cal destacar els relacionats amb el finançament i amb la propietat intel·lectual dels documents. Si obtenir un finançament adequat és el problema principal que fins ara ha aturat la posada en marxa de projectes a gran escala, tenir cura de no infrin-

gir la legislació de propietat intel·lectual també pot esdevenir un fre continuat. Pel que fa a les obres sota domini públic, no hi ha problemes perquè siguin digitalitzades i ofertes per qui ho vulgui, però per les obres amb drets de propietat intel·lectual vigents cal obtenir prèviament els corresponents drets de reproducció i difusió dels autors i editors. Cal indicar que dins d'aquest darrer grup hi ha un nombre molt important d'obres que es troben en una situació particular: legalment estan dins del període de protecció però se'n desconeix el titular dels drets o no se l'ha pogut localitzar, com seria el cas d'empreses editorials que ja no existeixen i d'autors o d'hereus seus il·localitzables. Aquestes «obres orfes» actualment no es poden digitalitzar, però diverses biblioteques i altres estaments culturals estan demanant que s'hi trobi una solució que permeti posar-les a l'abast dels ciutadans en suport digital.⁷

ELS PROJECTES DE GOOGLE

El projecte Google Print Library va ser presentat el 14 de desembre de 2004 i es proposa la digitalització dels fons de les biblioteques de University of Michigan, Harvard University, Stanford University, The New York Public Library i Oxford University. Google assumeix totes les despeses i també hi posa el personal i els mitjans tècnics; per la seva part les biblioteques faciliten els seus fons de llibres, totalment o en part: mentre que la University of Michigan té previst digitalitzar els 7 milions de llibres del seu fons, les universitats d'Stanford, Oxford, Harvard i la NYPL, només faran, de moment, proves pilot. El concepte de *prova pilot* és ben diferent del que podríem considerar en el nostre país: per a Stanford significa digitalitzar 2 milions de llibres (en deixen 6 per més endavant); per a Oxford és 1 milió de llibres de la Bodleian Library; i per a Harvard són 40.000 volums (després dels quals podria seguir la seva col·lecció sencera de 15 milions de llibres).

Malgrat que l'arribada a l'opinió pública de la notícia de l'acord va tenir un gran impacte, la seva gestació ja venia de lluny: Oxford, per exemple, ha reconegut que negociava amb Google des de l'any 2002.⁸ Els acords establerts entre les dues parts són confidencials i presenten variacions en cada cas,

7. La Library of Congress ha iniciat als Estats Units un procés en aquesta línia: *Report on orphan works: a report of the Register of Copyrights*. (Washington DC: Library of Congress. US Copyright Office, 2006) <<http://www.copyright.gov/orphan/orphan-report.pdf>> [Consulta: 7-2-2006].

8. Ronald Milne, «The Google Library Project at Oxford», *Internet Reference Services Quarterly*, vol. 10, núm. 3/4 (2005), p. 23-28 (p. 24).

però hom se'n pot fer una certa idea a partir de les informacions periodístiques i de la documentació penjada a la xarxa per Michigan.⁹ Es tracta d'un projecte de llarga durada, en el cas de Michigan s'estima que caldran sis anys d'escaneig i tres en el d'Oxford. Google agafa els llibres de les biblioteques per blocs i se'ls emporta en furgonetes cap a instal·lacions pròpies que ha construït fora dels campus universitaris (en el cas de les universitats de Michigan i d'Stanford) o en espais dins de la pròpia institució (en el cas de la NYPL i de les universitats d'Oxford i de Harvard). Totes les actuacions tècniques que porta a terme són secretes, atès que es troba en un mercat competitiu; se sap, però, que disposa de màquines de disseny propi per escanejar els llibres sense trencar l'enquadració. Google lliura a les biblioteques fitxers amb la imatge d'alta resolució de les pàgines escanejades i també fitxers amb el text obtingut per OCR. Aquests fitxers són propietat de les biblioteques, amb els quals esperen muntar sistemes de biblioteca digital integral: s'espera que en el futur una part significativa de la consulta bibliogràfica es faci en suport digital en aquestes institucions, deixant les obres originals en paper com a patrimoni a preservar, i també es preveu que puguin oferir per préstec interbibliotecari les versions digitals de les obres que siguin de domini públic. Google, per la seva banda, oferirà accés complert a les obres de domini públic i permetrà també cercar el text de les obres sota propietat intel·lectual, però només en mostrarà un petit fragment. Hom ha calculat que Google potser invertirà uns 200 milions de dòlars en el projecte.¹⁰

En el projecte només s'escanegen llibres i s'exceptuen els integrats en fons especials i de reserva. En el cas d'Oxford, de Harvard i de la NYPL s'especifica clarament que només es tractaran obres que siguin de domini públic; a Michigan es digitalitzaran tots els llibres, amb copyright i sense; a Stanford les informacions no són prou clares.

Google Print Library va comptar ben aviat amb dos tipus d'oposició: per infringir suposadament els drets dels editors i per consolidar la preeminència dels continguts en anglès. Pel que fa al primer problema, ben aviat es van sentir les queixes de grups com ara l'Association of Learned and Professional Society Publishers (ALPSP),¹¹ la Text and Academic Authors Association (TAA),¹² l'Authors Guild (representant de 8.000 autors nord-americans)¹³

9. <<http://www.lib.umich.edu/mdp/>> [Consulta: 7-2-2006].

10. «Microsoft scans British Library», *BBC News* (4 novembre 2005), <<http://news.bbc.co.uk/1/hi/technology/4402442.stm>> [Consulta: 7-2-2006].

11. <<http://www.alpsp.org/>> [Consulta: 7-2-2006].

12. <<http://www.taaonline.net/>> [Consulta: 7-2-2006].

13. <<http://www.authorsguild.org/>> [Consulta: 7-2-2006].

i l'Association of American Publishers (AAP).¹⁴ Es tracta de la representació de gairebé la totalitat de la indústria editorial i dels autors dels Estats Units.

Al llarg de la primavera i l'estiu de 2005 Google i les associacions oposidores van mantenir negociacions confidencials per arribar a algun tipus d'acord, però en no aconseguir-lo, les segones van decidir denunciar el conflicte davant dels tribunals de justícia. El 20 de setembre l'Authors Guild i tres autors individuals van demandar Google davant la New York Federal Court per infrinir el copyright.¹⁵ El 19 d'octubre Google va ser demandada davant de la U.S. District Court, Southern District of New York pels grups editorials membres de l'AAP McGraw-Hill, Pearson Education, Penguin Group, Simon & Schuster i John Wiley & Sons.¹⁶ El punt legal de divergència està en l'orientació «opt-out» del projecte en lloc de la de «opt-in» que reclamen els autors i editors; en termes més simples, Google vol negociar i obtenir els permisos editorials de les obres amb copyright després de digitalitzar-les en lloc de fer-ho abans d'iniciar-la. Més enllà de les queixes concretes, sembla que l'objectiu final de les editorials és aconseguir el tancament del Google Print Library Project i continuar amb l'anterior Google Print Publisher Program,¹⁷ un sistema d'acords bilaterals per mitjà del qual una determinada editorial permet que les obres que decideixi puguin ser parcialment cercables per Google a canvi de facilitar-ne la venda.

L'altre front d'oposició va venir d'Europa. Sota el lideratge del govern francès, es va expressar la por que Google Print refermés encara més la preeminència dels continguts en anglès a la xarxa. Aquesta recança estava justificada per l'origen dels fons (quatre biblioteques nord-americanes i una d'anglesa) que, lògicament, tindrien una composició lingüística i també intel·lectual esbiaixada respecte a la realitat mundial.¹⁸ En nom del manteniment de la diversitat cultural, es feia una crida a generar des d'Europa algun

14. <<http://www.publishers.org/>> [Consulta: 7-2-2006].

15. Comunicat oficial a: <http://www.authorsguild.org/news/sues_google_citing.htm>.

16. Comunicat oficial a: <<http://www.publishers.org/press/releases.cfm?PressReleaseArticleID=292>> [Consulta: 7-2-2006].

17. <<https://books.google.com/partner/>> [Consulta: 7-2-2006].

18. Aquest suposat biaix ha estat força contestat per un estudi d'OCLC: segons les seves estimacions, si el projecte de Google es porta finalment a terme sobre la totalitat dels fons de les cinc biblioteques participants, només el 49 % dels llibres digitalitzats seran en anglès. Brian Lavoie; Lynn Sili-pigni Connaway; Lorcan Dempsey, «Anatomy of agregate collections: the example of Google Print for Libraries», *D-Lib Magazine*, vol. 11, núm. 9 (2005) <<http://www.dlib.org/dlib/sep-tember05/lavoie/09lavoie.html>> [Consulta: 7-2-2006].

tipus de projecte d'un abast equivalent que compensés els continguts que abocaria Google a la xarxa.

Bloquejada pel conflicte judicial i preocupada pels problemes d'imatge que li començaven a suposar les crítiques, Google va emprendre al llarg de la tardor una certa reorientació del projecte. A partir del 17 de novembre de 2005 Google Print es va passar a dir Google Book Search i va començar presentant-se de manera més clara, com un servei que «només» volia facilitar la cerca de llibres, com si es tractés del catàleg d'una biblioteca amb tecnologia actualitzada, perquè s'oferiria una referència bibliogràfica dels llibres a més de breus fragments del text, en el cas de les obres sota propietat intel·lectual, o el text complert, si eren obres de domini públic.

Gairebé al mateix temps es van presentar versions no angleses del servei, com l'espanyola <<http://books.google.es>>. Cal fixar-se que aquestes versions no són en realitat el resultat de projectes de digitalització retrospectiva, sinó d'acords puntuals amb alguns dels principals editors de cada país: en el cas d'Espanya, amb els grups editorials Anaya i Planeta.

ALTRES PROJECTES INTERNACIONALS

El projecte de Google ha tingut un gran ressò als mitjans de comunicació i, de moment, ja ha tingut de positiu revalorar aquest tipus d'activitats. Com a resultat, les empreses competidores de Google han fet públiques iniciatives similars i també s'ha tornat a parlar de projectes més antics ja en curs. A continuació fem un repàs de les característiques de tots plegats.

Open Content Alliance (OCA)

L'Open Content Alliance (OCA)¹⁹ és una associació comandada per l'Internet Archive i amb la participació, entre altres, de la Smithsonian Institution, Columbia University, Toronto University, National Archives in the United Kingdom, O'Reilly Media and Prelinger Archives, University of California (California Digital Library), Johns Hopkins University, Research Libraries Group (RLG), Yahoo i Adobe. Es van presentar el 3 d'octubre de 2005 amb l'objectiu de crear un arxiu permanent de continguts textuals digitalitzats i multimèdia. Poc després, el 25 d'octubre s'hi va afegir Microsoft per mitjà del seu portal MSN.

19 <<http://www.opencontentalliance.org/>> [Consulta: 7-2-2006].

La definició del projecte és força clara. Unes institucions (biblioteques i arxius) proporcionen els fons a digitalitzar; Adobe, Hewlett Packard i Xerox bàsicament aportaran llicències de programari, Microsoft i Yahoo finançaràn l'operació i Internet Archive coordinarà els aspectes tècnics, gràcies a la seva experiència en la preservació del web. Els resultats, potser sota formes diferents, es podran consultar per tres vies: al portal de Yahoo, al portal MSN de Microsoft i al propi web de l'OCA.

Aquest projecte ha estat força transparent respecte a les seves característiques tècniques.²⁰ La tecnologia que s'emprarà és la digitalització amb màquines d'escanejar llibres desenvolupades per Internet Archive i Kirtas, i el programari de digitalització serà Scribe, de codi obert.²¹ Aquesta combinació permetrà escanejar les pàgines a un cost unitari de 10 cèntims de dòlar.

Cal destacar la participació de Microsoft, que ha promès l'aportació inicial de cinc milions de dòlars per escanejar 150.000 llibres al llarg de l'any 2006. Amb els continguts que obtingui, preveu obrir el nou servei MSN Book Search, un clar competidor de l'oferta de llibres de Google. Mentrestant, l'Internet Archive ja ha obert l'Open Library²² com a seu del futur portal de llibres digitalitzats i com a demostració inicial del servei que es pensa oferir. De moment només s'hi presenten deu llibres de mostra, proporcionats per la University of California, però els resultats ja són engrescadors: els llibres no es presenten en forma d'imatges separades per pàgines, com fins ara ha estat habitual en els projectes de digitalització, sinó que s'ha aplicat una il·lusió òptica que simula el pas de les pàgines per l'usuari; és una solució similar al *turning the pages* de la British Library,²³ però més eficient tècnicament. Els llibres no tan sols es podran llegir, sinó que també es podran descarregar en format PDF i, fins i tot, se'n podrà escoltar la lectura gràcies a la transmissió del so per *streaming*.

The Universal Library

The Universal Library²⁴ es un projecte d'escaneig de llibres procedents de biblioteques promogut per la Carnegie Mellon University. En deu anys

20. Brewster Kahle, *The Open Library*, <<http://www.openlibrary.org/details/openlibrary>> [Consulta: 7-2-2006].

21. <<http://sourceforge.net/projects/scribesw/>> [Consulta: 7-2-2006].

22. <<http://www.openlibrary.org/>> [Consulta: 7-2-2006].

23. <http://www.armadillosystems.com/ttp_commercial/home.htm> [Consulta: 7-2-2006].

24. <<http://tera-3.ul.cs.cmu.edu/>> [Consulta: 7-2-2006].

volen arribar a oferir deu milions de llibres escanejats, però ara el seu objectiu és la Million Book Digital Library, amb un milió de llibres, que pensen tenir completada el 2007. El finançament procedeix de subvencions de la National Science Foundation (EUA), d'aportacions en espècie (personal i instal·lacions) dels governs de la Xina i l'Índia, l'hostatge digital per l'Internet Archive i el lliurament gratuït de metadades bibliogràfiques per OCLC.²⁵

Aquest projecte no es basa en l'aplicació de sistemes avançats d'automatització, com els projectes de Google i l'OCA, sinó en l'ús de mà d'obra barata de països emergents. Dels 600.000 llibres fins ara escanejats, 420.000 ho han estat a la Xina, 170.000 a l'Índia i uns 20.000 a Egipte. Per fer-se càrrec de la complicada logística del projecte, cal remarcar que milers de llibres estan sent enviats per avió dels Estats Units cap als centres de digitalització d'Àsia i retornats al cap de pocs dies.

La participació de la Xina i l'Índia, per mitjà dels seus governs, d'organismes de recerca i d'universitats es tradueix en un gran pes del xinès i de continguts propis d'aquests països. Els resultats del projecte ja són parcialment visibles per mitjà de tres portals: una entrada als Estats Units <<http://tera-3.ul.cs.cmu.edu/>>, una entrada a la Xina <<http://www.ulib.org.cn>> i la Digital Library of India <<http://dli.iit.ac.in/>>. La interfície de consulta és força austera i alguns dels resultats consultables no són de bona qualitat tècnica.

Biblioteca Digital Europea

L'anunci del projecte Google Print Library va provocar una gran reacció contrària a França. Cal recordar que aquest país ha estat durant anys dins d'Europa i davant de tot el món el màxim defensor de l'excepció cultural, de la idea que cal protegir les diverses cultures contra les pràctiques purament comercials, que tendeixen a promoure una cultura d'abast global, generalment en anglès i només sota estrictes regles de rendiment econòmic.

El moviment es va iniciar el 22 de gener de 2005 amb un article del president de la Biblioteca Nacional de França²⁶ i ben aviat va obtenir el suport

25. *Frequently asked questions about the Million Book Project*, <http://www.library.cmu.edu/Libraries/MBP_FAQ.html> [Consulta: 7-2-2006].

26. Jean-Noël Jeanneney, «Quand Google défie l'Europe», *Le Monde* (22 janvier 2005). Disponible a: <<http://mail.wikipedia.org/pipermail/wikifr-l/2005-January/002374.html>> [Consulta: 7-2-2006].

de les màximes autoritats franceses. La configuració política definitiva va prendre la forma d'una carta formal enviada el 28 d'abril a Jean Claude Juncker, primer ministre luxemburguès i en aquells moments president del Consell Europeu, en la qual es demanava la creació d'una biblioteca digital europea; signaven la petició el president de la República Francesa, el president de Polònia, el canceller de la República Federal d'Alemanya, el president del Consell de Ministres d'Itàlia, el president del Govern espanyol i el primer ministre d'Hongria.²⁷ El projecte no ha arribat a entrar dins les actuacions de la Unió Europea, sinó que s'està desenvolupant com una actuació paral·lela sota la coordinació francesa, que es va materialitzar el juliol amb la creació d'un comitè d'estudi constituït en exclusiva per representants de diversos ministeris francesos.²⁸ Al nucli dur format per sis països, s'hi ha afegit després la resta amb l'absència, comprensible, del Regne Unit.

La informació tècnica fins ara proporcionada és ben minsa. Els objectius declarats són digitalitzar de 600.000 a un milió de llibres l'any, però només França ha avançat que hi pensa invertir entre vuit i quinze milions d'euros anuals per digitalitzar de 150.000 a 200.000 llibres l'any; de la resta de països no se'n sap gairebé res. Per tot plegat, cal arribar a la conclusió que no ens trobem davant d'un veritable projecte d'abast europeu, sinó de la suma d'iniciatives nacionals i de, com a molt, l'acceleració d'alguns projectes anteriors.

Per últim assenyalar que el 9 de novembre de 2005 la Comissió de Política Cultural del Parlament de Catalunya va adoptar una resolució per la qual instava el Govern de la Generalitat «a dur a terme davant del Govern de l'Estat espanyol i de les Comissions corresponents de la Unió Europea les actuacions institucionals i polítiques adients per tal que el patrimoni literari i científic català formi part de la futura Biblioteca Digital Europea».²⁹

27. Text complet a: <http://www.elysee.fr/elysee/francais/interventions/lettres_et_messages/2005/avril/message_a_propos_de_la_creation_d_une_bibliotheque_numerique_europeenne.29633.html> [Consulta: 7-2-2006].

28. <<http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=MCCX0508574D#>> [Consulta: 7-2-2006].

29. BOPC, 250 (28/11/2005), p. 6. <<http://www.parlament-cat.net/activitat/bopc/07b250.pdf>> [Consulta: 7-2-2006].

Project Gutenberg

El Project Gutenberg³⁰ és possiblement el projecte de digitalització més antic del món. Es va iniciar l'any 1971 amb l'objectiu de popularitzar l'accés a les principals obres del saber humà, amb independència de la llengua en què es van imprimir. Actualment té uns 17.000 llibres digitalitzats, tots a partir de llibres de domini públic. Es basa en la feina de voluntaris de tot el món, que s'encarreguen de transcriure les diverses obres.

Atès que el projecte està focalitzat en els continguts, en les obres dels autors, i no pas en com es van presentar en una edició determinada, el Project Gutenberg no té fitxers gràfics amb les imatges de pàgines de llibres, sinó el text transcrit i ofert en diversos formats. Durant molts anys el text només va estar disponible en format ASCII, però actualment també es pot trobar en RTF, HTML o altres, que poden variar d'una obra a altra. Els seus fons estan oberts, sense ànim de lucre, i es poden incorporar a altres projectes de biblioteca digital.

World Digital Library (WDL)

La World Digital Library (WDL) actualment només és un projecte de la Library of Congress sense massa concreció.³¹ S'ha presentat com un portal per la consulta de les versions digitals dels materials de tipus internacional que formen part de les col·leccions de la biblioteca. Així es vol completar el paper que el portal American Memory³² ja fa respecte als materials digitals propis dels Estats Units.

La WDL es crearà a partir les col·leccions que ara ja es poden consultar al Global Gateway,³³ que són el resultat de projectes pilots iniciats a redós del Grup dels Set (G-7) i d'iniciatives posteriors de la mateixa Library of Congress. La WDL serà de consulta pública i, en part, es pensa finançar amb donacions d'empreses privades; la primera va ser, significativament, de Google, amb l'aportació el novembre de 2005 de tres milions de dòlars.³⁴ Tampoc es descarta que més endavant pugui obrir-se a col·laboracions d'organismes internacionals o d'altres països.

30. <<http://www.gutenberg.org/>> [Consulta: 7-2-2006].

31. La idea es va llançar el juny de 2005, vegeu: <http://www.loc.gov/about/welcome/speeches/wdl/wdl_6-6-05.html> [Consulta: 7-2-2006].

32. <<http://memory.loc.gov/ammem/>> [Consulta: 7-2-2006].

33. <http://international.loc.gov/intldl/find/digital_collections.html> [Consulta: 7-2-2006].

34. <<http://www.loc.gov/today/pr/2005/05-250.html>> [Consulta: 7-2-2006].

Amazon

Amazon³⁵ és una botiga de venda en línia de llibres i de molts altres articles que sempre s'ha caracteritzat per la creativitat a l'hora d'incentivar els usuaris amb nous reclams. Ja fa uns tres anys va començar a oferir de manera gratuïta el contingut d'una part dels llibres que ven; després d'arribar a acords amb els editors, integra el text dels llibres en el seu cercador, permetent que els usuaris en facin ús i en puguin fer una visualització limitada. Segons l'empresa això ha fet augmentar les vendes dels llibres que incorporen aquesta funció que, d'altra banda, no és molt diferent al fullejar que fan els clients de les llibreries tradicionals.

El novembre passat va anunciar que estan preparant nous desenvolupaments, encara no disponibles pel públic. Són: Amazon Pages, que permetrà comprar i llegir en línia només la part que es desitgi d'un llibre, sigui un capítol o una pàgina concreta; i Amazon Upgrade, que permetrà mitjançant el pagament d'un suplement accedir a la versió digital d'un llibre que prèviament s'hagi comprat en paper.³⁶

The European Library (TEL)

The European Library³⁷ vol actuar com a portal comú de les col·leccions digitals de 43 biblioteques nacionals europees, si bé de moment només ofereix els fons de quinze països; la coordinació tècnica va a càrrec de la Koninklijke Bibliotheek, dels Països Baixos. És el resultat de la fusió de diversos projectes pilot anteriors, com la Bibliotheca Universalis³⁸ i el lloc web Gabriel, que fins l'estiu de 2005 actuava com a portal d'accés a les biblioteques nacionals europees.

35. <<http://www.amazon.com/>> [Consulta: 7-2-2006].

36. Comunicat oficial a: <<http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=778248&highlight=>>> [Consulta: 7-2-2006].

37. <<http://www.theeuropeanlibrary.org/>> [Consulta: 7-2-2006].

38. La Bibliotheca Universalis va ser un dels projectes pilot promoguts pel grup del G-7 dins del seu pla Global Information Society, desenvolupat entre el 1995 i el 2002, any en què es va obrir el lloc web per la consulta pública. Inicialment volia integrar materials de diversos països i cultures sota el títol *Exchanges between people*, mostrant relacions més enllà de les fronteres polítiques, però al final els problemes tècnics i d'altre tipus van reduir Bibliotheca Universalis a un simple directori que donava accés a diverses col·leccions digitals sense massa relació entre elles; hi participen les biblioteques nacionals d'Alemanya, Austràlia, Bèlgica, Canadà, Espanya, Estats Units d'Amèrica, França, Itàlia, Japó, Països Baixos, Portugal, Suïssa, Regne Unit i la República Txeca.

La principal raó de ser de The European Library és actuar com a sistema d'accés comú a les col·leccions digitals de les biblioteques participants. TEL no executa projectes propis ni disposa de finançament per fer-ho, sinó que interroga les col·leccions nacionals i hi dóna accés en els seus llocs d'origen. Entre els materials disponibles ja hi ha llibres digitalitzats, molts dels quals procedeixen de Gallica,³⁹ la biblioteca digital de la Bibliothèque Nationale de France. Potser com a resultat d'aquest origen heterogeni, resulta difícil consultar els continguts per categories, com llibre digital, i després són oferts amb formats diferents i amb interfícies de consulta pròpies de cada biblioteca.

British Library

La British Library va ser una de les primeres institucions del món en estudiar la digitalització de llibres. Una de les seves principals aportacions va ser el sistema *turning the pages*,⁴⁰ que simula el pas de les pàgines d'un llibre digitalitzat en sistemes de consulta en mode quiosc o en CD-ROM. Actualment és una tecnologia que ha estat superada⁴¹ però que continua mostrant la seva espectacularitat si hom la utilitza en llocs com la British Library o el British Museum.

La British Library té un programa propi de digitalització a partir d'una selecció d'obres. En el context de la lluita entre els grans cercadors d'Internet per oferir continguts exclusius, el novembre de 2005 va aconseguir que Microsoft li financés, en una primera fase, la digitalització de 25 milions de pàgines. Els resultats seran visibles per mitjà de la British Library i dels portals de Microsoft.⁴²

REFLEXIONS FINALS

A les pàgines anteriors hem resseguit les característiques dels projectes que per volum o objectius declarats tenen un abast «universal». Cal no oblidar que paral·lelament es porten a terme molts altres projectes de menor escala que es desenvolupen en països o biblioteques determinats. Pel que fa a allò que ens és més proper, cal deixar constància que, malauradament, a Espanya no hi ha

39. <<http://gallica.bnf.fr/>> [Consulta: 7-2-2006].

40. <<http://www.bl.uk/onlinegallery/ttp/tpbooks.html>> [Consulta: 7-2-2006].

41. S. K. Card; et al. «3Book: a scalable 3D virtual book». En: *Extended Abstracts of the ACM International Conference on Human Factors in Computing Systems*. (Viena: ACM, 2004), p. 1095-1098.

42. <<http://news.bbc.co.uk/1/hi/technology/4402442.stm>> [Consulta: 7-2-2006].

gaire projectes consolidats,⁴³ amb l'excepció que tothom coneix de la Biblioteca Virtual Miguel de Cervantes.⁴⁴ La Biblioteca Nacional de Madrid no té un pla extens de digitalització de llibres i el gener d'enguany només oferia la consulta de 405 títols en suport digital;⁴⁵ a la Biblioteca de Catalunya la situació és similar, amb uns pocs centenars de llibres digitalitzats. Ambdues institucions fins ara s'han recolzat en la col·laboració amb la Biblioteca Virtual Miguel de Cervantes, que ha assumit els costos i els aspectes tècnics de la digitalització d'obres escollides, i la creació de portals específics d'entrada.⁴⁶

Si a escala local el panorama és decebedor, a escala global la digitalització dels llibres de les biblioteques s'ha posat de moda gràcies als interessos comercials: la lluita entre Google, Yahoo i Microsoft per ser el principal distribuïdor de continguts a Internet. Les biblioteques han vist com se'n revalorava la funció de conservació del patrimoni documental i com la feina acumulada al llarg d'anys i de segles ara continua sent necessària. Els mitjans digitals fan que puguin desplegar els seus serveis més enllà del seu radi d'acció territorial i, si cal, arribar a tot el món. Algunes grans biblioteques, és cert, les més ben preparades, se n'han pogut aprofitar i aviat disposaran d'uns fons pràcticament digitalitzats que els permetran oferir nous serveis automatitzats.

El moment també presenta un costat obscur. Molts professionals de la informació senten amenaçada la funció de les institucions per les quals treballen i això ha fet que s'obri un debat en el qual, de moment, trobem moltes preguntes però poques respostes concretes.⁴⁷ Fins a quin punt les biblioteques estan «regalant» el seu patrimoni a canvi d'uns fitxers que potser són de pitjor qualitat que els que es queden les multinacionals? Seran necessàries les gran biblioteques de reserva si bona part dels seus continguts seran accessibles a la xarxa? O potser no es tracta de papers o de bits, sinó de biblioteques gratuïtes (però aparentment tancades) enfront de serveis digitals de pagament, sota diverses formes (però aparentment oberts)? Certament, la millora en ac-

43. Hom pot trobar un inventari força actualitzat a l'apèndix de l'obra *Directrices para proyectos de digitalización de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos*. (Madrid: Ministerio de Cultura, 2005), p. 169-183. Disponible a <http://travesia.mcu.es/documentos/pautas_digitalizacion.pdf> [Consulta: 7-2-2006].

44. <<http://www.cervantesvirtual.com/>> [Consulta: 7-2-2006].

45. <<http://www.bne.es/cgi-bin/wsirtex?FOR=WBNCNS4>> [Consulta: 7-2-2006].

46. <<http://www.cervantesvirtual.com/portal/bne/>> i <<http://www.cervantesvirtual.com/portal/BC/>> [Consulta: 7-2-2006].

47. Un exemple d'aquesta preocupació el tenim en el número monogràfic «Libraries and Google» de la revista *Internet Reference Services Quarterly*, vol. 10, núm. 3/4 (2005). També John Maxymuk, «13 Ways of looking at Googlebrary», *The Bottom Line: Managing Library Finances*, vol. 18, núm. 4 (2005), p. 197-201.

cessibilitat i disponibilitat de les obres es notarà, però la mediatització dels cercadors i de les necessitats empresarials també. Per pensar-hi només cal navegar una mica per Amazon, l'autoanomenada «major llibreria del món», per adonar-se que ofereix centenars de milers de llibres, però sempre intenta vendre un nombre concret d'aquests, que sempre són del mateix tipus.

Oberta, global, mundial són alguns dels adjectius que s'adjudiquen les noves biblioteques digitals. Aquí les xifres ens enganyen: parlen de milions de llibres a digitalitzar com si per això sols ja tinguessin un caràcter universal. Però les biblioteques, fins i tot les més grans, creen les seves col·leccions pensant en usuaris determinats, amb necessitats determinades i en un context determinat; el biaix, per tant, existeix i és bo. El problema és dels *altres*, els que no vivim a Califòrnia, a Nova York, a Oxford o no hem nascut d'uns pares que parlessin en anglès.

Si el fals universalisme pot ser un problema per a les llengües i les cultures minoritàries (o potser per totes les no anglosaxones), el minifundisme i el localisme poden portar a mirar-se el melic sense arribar a salvar la cultura pròpia. De poc serveix en un món globalitzat omplir portals amb uns centenars de llibres locals digitalitzats i oferts amb pocs mitjans.

També es pot parlar, amb una mica de nostàlgia, dels temps BG, «before Google», però aquests ja no tornaran, ara el món de la informació ha canviat amb una xarxa Internet que permet resoldre multitud de necessitats informatives en pocs de segons.⁴⁸ Michael Gorman, president de l'American Library Association (ALA), ha alertat no tant sobre l'impacte d'aquestes iniciatives en les biblioteques, sinó sobre l'ús que en fan els usuaris: Google atomitza el coneixement perquè el serveix en petites dosis, pot reduir la lectura dels llibres a només la dels petits paràgrafs que oferirà, privant als lectors desprevenguts del sentit real que només podrien obtenir amb l'obra sencera.⁴⁹ La banalització dels serveis i el seu empaquetat per a públics cada cop més amplis ha portat algunes persones a parlar de manera despectiva de la *macdonalització* de les biblioteques,⁵⁰ però aquest potser és el preu que cal pagar per democratitzar més els serveis i ampliar-ne l'oferta. Les biblioteques han de reconèixer aquesta nova realitat en la qual han aparegut competidors cada cop més grans que, a causa dels seus interessos, han arribat per quedar-s'hi, amb la nostra col·laboració o sense.

48. Rick Anderson, «The (Uncertain) Future of Libraries in a Google World: Sounding an Alarm», *Internet Reference Services Quarterly*, vol. 10, núm. 3/4 (2005), p. 29-36.

49. Andreas von Bubnoff, «The real death of print», *Nature*, vol. 438, núm. 7068 (1 December 2005), p. 550-552 (p. 550).

50. Brian Quinn, «The McDonaldization of Academic Libraries?», *College and Research Libraries*, vol. 61, núm. 3 (May 2000). p. 248-261.