

Preservació de recursos d'informació digital

Alice Keefer Riva
Núria Gallart Marsillas

PID_00187101

Material docent de la UOC

**Alice Keefer Riva**

Llicenciada en Filologia Hispànica per la Pennsylvania State University i Màster en Biblioteconomia pel Simmons College (Boston, USA). Abans de traslladar-se a Espanya, va treballar en diferents biblioteques: la Library of Congress, la de la Pennsylvania State University, la de l'Estat de Massachusetts i la de l'Organització dels Estats Americans, especialitzant-se sobretot en aspectes relacionats amb les publicacions en sèrie. A Barcelona, des de la seva arribada el 1984, ha estat directora de biblioteca de l'Institut d'Estudis Nord-americans i posteriorment, durant 10 anys, Directora de Serveis Internacionals de l'empresa de serveis DOC6. Des de 1996 és professora associada de la Facultat de Biblioteconomia i Documentació de la Universitat de Barcelona on imparteix assignatures sobre diferents aspectes de l'adquisició, gestió i preservació de recursos electrònics, especialment de publicacions en sèrie. És consultora en l'assignatura "Preservació de recursos digitals" de la UOC des de l'any 2000. És autora de nombrosos articles sobre temes relacionats amb la gestió i preservació dels recursos digitals a les biblioteques i, sovint, imparteix cursos sobre aquests mateixos temes en diferents ciutats espanyoles.

**Núria Gallart Marsillas**

Llicenciada en Documentació. Actualment té a càrrec seu la direcció de la Biblioteca Universitària de Sabadell de la Universitat Autònoma de Barcelona que dona servei a les Escoles Universitàries d'Estudis Empresarials i d'Informàtica. Col·labora amb la Universitat Oberta de Catalunya com a consultora dels Estudis de Ciències de la Informació i de la Comunicació. Entre 1993 i 2004 va ser Cap de projectes del Servei de Biblioteques de la UAB i com a tal va coordinar els equips locals que van intervenir en diversos projectes finançats pel Programa europeu Telemàtica per a Biblioteques. Membre del Grup BPI –Biblioteques i Propietat Intel·lectual– de FESABID des de la seva fundació el 1996 fins a 2002.

La revisió d'aquest material docent ha estat coordinada pel professor: Àlex López Borrull (2012)

Segona edició: febrer 2012
 © Alice Keefer Riva, Núria Gallart Marsillas
 Tots els drets reservats
 © d'aquesta edició, FUOC, 2012
 Av. Tibidabo, 39-43, 08035 Barcelona
 Disseny: Manel Andreu
 Realització editorial: Eureka Media, SL
 Dipòsit legal: B-5.149-2012

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit dels titulars del copyright.

Introducció

Que els recursos digitals ja estan immersos en totes les esferes de la nostra societat és un fet innegable i inevitable. En l'última dècada, els avenços de la tecnologia informàtica i de les telecomunicacions han facilitat la creació, distribució i ús d'informació digital. Una de les conseqüències és que els avantatges (amb alguns inconvenients) han arribat a totes les capes de la societat als països desenvolupats.

Un passeig per alguns titulars recents és un bon punt de partida per a reflexionar sobre la presència d'eines i informació digital i els seus possibles efectes:

- *Google digitalizará más de 300.000 libros de cinco bibliotecas catalanas* (LaVanguardia.es, 10.01.07)
- *Llega Windows Vista, el nuevo sistema operativo de Microsoft* (Reuters España, 30.01.07)
- *Microsoft lanza Windows Mobile 6, la última versión de su sistema para teléfonos móviles* (PC Actual, 12/02/07)
- *Windows 2010, el sucesor de Vista [a finales del 2009]* (Noticiasdot.com, 13.02.07)
- *Microsoft saca al mercado español su DVD de alta definición* (EFE/Invertia, 15/2/07)
- *Google venderá por Internet paquetes de software* (Yahoo España Noticias, 22.02.07)

A més d'anunciar nous productes i serveis, aquestes notícies –aparegudes durant un període de sis setmanes– deixen entreveure un mercat caracteritzat per canvis constants, en el qual cada novetat condueix al creixement, també imparable, de recursos d'informació digital.

De la mateixa manera que la immersió tecnològica ha arribat a totes les esferes de la vida pública i privada, també es fan sentir els efectes d'aquests canvis. Mentre que hi ha qui només veurà els avantatges que representen les millores de maquinari, de programari i dels recursos creats per la combinació de tots dos, d'altres se sentiran vulnerables davant d'aquests mateixos canvis. Per exemple, moltes organitzacions temen que els programes i la informació de l'entorn informàtic anterior resultin inoperables o inaccessibles per l'evolució

tecnològica. Alhora, se senten impotents davant de l'obligació imposada pel mercat d'adaptar-se ineludiblement a les innovacions, que sovint comporta un gran cost econòmic i, de vegades, organitzacional.

Entre les organitzacions que en resulten afectades, n'hi ha que s'encaren a un doble repte; es tracta de les institucions de "memòria cultural"; és a dir, de biblioteques, arxius i museus. Aquestes sempre s'han responsabilitzat de la preservació de material tradicional; per això, a més de tenir el repte d'adaptar les seves operacions quotidianes als canvis, també es pregunten com podran gestionar i preservar la quantitat ingent d'informació creada: la seva selecció, descripció, organització i preservació. No solament s'han de preocupar de la informació de creació pròpia, sinó dels recursos generats per altres organitzacions i que, finalment, arribaran a les biblioteques, els arxius i els museus per a la seva custòdia perpètua.

1. Internet i el World Wide Web

La liberalització de l'accés a Internet al començament dels noranta va marcar l'inici del procés d'eliminació dels obstacles per a l'accés a la informació imposats per les parets físiques de la biblioteca. Les funcions inicials d'Internet –correu electrònic, transferència de fitxers i accés remot (Telnet)– van permetre el desenvolupament de serveis que van obrir grans expectatives. Avui els Gophers, WAIS i altres artefactes d'aquells anys poden semblar tremendament primitius, però en el seu dia representaven per als usuaris avenços importants per a consultar fonts d'informació llunyanes, i fins llavors inaccessibles.

A mitjan dècada dels noranta, va aparèixer el World Wide Web (WWW) amb el seu efecte arrasador: una nova manera –àgil, fàcil, econòmica– de distribuir informació i d'accedir-hi. La utilització dels instruments d'Internet previs al WWW es limitava a especialistes, a causa de factors com el tipus de material disponible a la xarxa (també especialista), la poca fiabilitat i l'alt cost de les telecomunicacions, i la complexitat tant dels llenguatges de recuperació com del maneig de les aplicacions.

Els lectors d'aquestes línies seran testimonis de l'allau de serveis d'informació ja disponibles mitjançant el WWW: gratuïts i de pagament, públics i privats, lícits i il·lícits. Les tecnologies i les aplicacions es multipliquen i faciliten encara més l'intercanvi d'idees i dades i, alhora, transformen la manera de comunicar-se les persones entre elles, mitjançant *wikis*, P2P, blocs, fotoblocs i altres eines informàtiques.

2. La preservació digital i les biblioteques: els primers anys

Òbviament, les biblioteques no han estat immunes a aquests canvis. Van entrar en el nou segle amb un peu posat fermament al món tradicional dels materials analògics i l'altre tocant lleugerament al món digital; en aquell moment, l'any 2000, feia aproximadament una dècada que disposaven de catà-

legs automatitzats i de material documental en suports digitals, principalment en discos òptics. Durant els primers anys, a més d'adoptar les noves eines per al seu propi ús i d'adquirir i gestionar recursos d'informació digital per als seus usuaris, van oferir a molts d'ells el primer contacte amb les noves tecnologies, mitjançant els lectors de CD i DVD i els terminals d'accés a Internet que van posar a disposició del públic.

El col·lectiu de biblioteques, que sempre havia assegurat la preservació dels materials bibliogràfics de segles enrere, va observar amb gran preocupació l'augment imparable de recursos d'informació digital en els seus fons, ja que no tenien els coneixements tècnics ni els mitjans econòmics i materials per a assumir la responsabilitat de la seva preservació indefinida i el seu accés en el futur.

Mentre que alguns professionals poden sentir-se amenaçats per aquests canvis i pel futur incert del fons digital, d'altres veuen les oportunitats que la transició els brinda: l'aplicació dels coneixements existents als nous recursos i l'adquisició, sobre la marxa, de nous coneixements. Si aquests dos grups representen els extrems, al mig hi ha molts professionals –possiblement la majoria– disposats a afrontar els canvis i en espera de saber què hauran de fer i la manera de fer-ho, per a poder donar una resposta adequada als nous reptes.

Un dels primers articles publicats sobre la inestabilitat de la informació digital va ser "¿Son perdurables los documentos digitales?" aparegut a *Investigación y Ciencia* el març del 1995. Jeff Rothenberg hi especula sobre el que haurien de fer els seus néts si deixés les instruccions per a trobar la seva hipotètica fortuna en un CD-ROM. Com a conclusió, Rothenberg contesta la pregunta formulada en el títol de la manera següent: [Si el CD inclogués, a més de la clau] tota la base informàtica requerida juntament amb especificacions completes i fàcils de desxifrar concernents a l'equip físic necessari, potser... [s'aconsegueixi] generar un emulador capaç d'executar els programes originals encarregats de mostrar el document.

Paral·lelament a aquesta primera crida d'atenció, al món bibliotecari s'obrien diversos fronts per a encarar el problema de la permanència dels documents digitals i de les possibilitats d'utilitzar-los en el futur. En els paràgrafs següents es destaquen algunes de les institucions que van protagonitzar les primeres iniciatives: algunes van ser liderades per les universitats o per les institucions que les representaven, a la Gran Bretanya i als Estats Units; d'altres van ser dutes a terme per les biblioteques nacionals australiana, holandesa i sueca.

2.1. Iniciatives universitàries

El 1996, gairebé simultàniament, es van elaborar dos informes –als Estats Units i a la Gran Bretanya respectivament– que van servir per a llançar una sèrie d'estudis i projectes posteriors:

Estats Units

Es va publicar l'informe *Preserving digital information*, subvencionat pel Research Libraries Group (RLG) –un consorci de grans biblioteques d'investigació, majoritàriament les universitàries– i la Commission on Preservation and Access (CPA). Aquest document representava el primer pas cap a la definició del problema i proposava recomanacions per a respondre-hi. Hi van participar diferents sectors i institucions molt prestigioses:

- **Sector universitari:** Columbia, Johns Hopkins, Universitat de Califòrnia, Universitat de Michigan, Yale.
- **Sector públic:** Library of Congress, National Library of Agriculture, National Archives and Records Administration.
- **Sector informàtic:** IBM, Bell Communications Research.
- **Sector editorial:** American Physics Society, Elsevier.

Una de les recomanacions d'aquest informe va ser la realització d'un estudi sobre la vertadera situació a les biblioteques d'investigació. Aquest es va dur a terme el 1998 i es va publicar amb el títol *Digital preservation needs and requirements in RLG member Institutions* (Hedstrom i Montgomery, 1998).

El segon document publicat també el 1996 va ser *Preservation in the digital world*, de Paul Conway, director de preservació a la biblioteca de la Universitat de Yale; aquest suggereix els canvis necessaris per a adaptar les operacions tradicionals de preservació a la nova situació i enfrontar-se al repte de la preservació dels materials digitals (Conway, 1996).

Un resultat d'aquestes primeres iniciatives va ser la creació del Council on Library and Information Resources (CLIR) que, a més, patrocinava la Digital Library Federation (DLF), consorci de biblioteques i agències bibliotecàries pioneres en l'ús de la tecnologia de la informació per al desenvolupament de les seves col·leccions i serveis. La DLF pretén identificar normes i "best practices", coordinar l'R+D en l'àrea i promoure projectes i serveis que una sola biblioteca no tindria possibilitats de fer. Un bon indicador del protagonisme de les universitats en el DLF és el fet que dels seus trenta membres fundadors, vint-i-sis eren biblioteques universitàries. Els altres quatre membres eren la Library of Congress, la New York Public Library, els National Archives and Records Administration i el patrocinador, el CLIR.

Gran Bretanya

A diferència dels nord-americans, els primers esforços dels quals procedien de consorcis de biblioteques universitàries, els britànics van emprendre iniciatives que estaven finançades i coordinades per entitats nacionals i, molt espe-

cialment, pel Joint Information Systems Committee (JISC), el comitè consultiu de les entitats públiques responsables del finançament de l'ensenyament superior al Regne Unit.

El 1995 el JISC i la British Library van organitzar les primeres jornades sobre la preservació a llarg termini dels materials electrònics, les conclusions de les quals es van publicar, el 1996, en l'informe *Long term preservation of electronic materials*. Aquest primer esforç va servir de guia per a nombrosos projectes posteriors finançats per l'Administració britànica.

L'any anterior a aquesta reunió, el JISC havia finançat una sèrie de projectes de digitalització de fons antics i valuosos per a millorar-ne l'accés, mitjançant el programa *eLib: The Electronic Libraries Programme* (1994-1997).

A causa de la preocupació despertada per la perdurabilitat d'aquests recursos digitals de nou generats i pagats per les arques públiques, el JISC va finançar un altre projecte, el CEDARS, dedicat a la problemàtica de la preservació digital (Beagrie, 2001).

Aquest nou projecte, de gran impacte internacional que va durar del 1998 al 2002, va disposar de la participació de les universitats de Leeds, Cambridge i Oxford, i va fer una de les primeres aproximacions a l'estudi dels costos que sens dubte impactarien sobre les estratègies de preservació (Granger i altres, 2000). Van ser molt destacables les guies que donaven a conèixer els resultats del projecte i que tenien com a objectiu els temes següents: drets de propietat intel·lectual, metadades per a la preservació, gestió de la col·lecció, estratègies de preservació i prototip d'arxiu digital.

CAMILEON, un altre conegut projecte, mostra l'alt nivell de col·laboració entre les biblioteques anglosaxones, ja que va ser realitzat conjuntament per la Universitat de Leeds a Anglaterra i la Universitat de Michigan als Estats Units. L'objectiu del projecte era investigar l'emulació com a estratègia de preservació digital, una estratègia promoguda per Rothenberg en el seu informe *Avoiding technological quicksand* o Com es poden evitar els sorramolls.

El seu autor va sostenir que l'emulació era l'única estratègia teòricament capaç de garantir que es preservés el *look and feel* original d'un recurs –alguna cosa així com una combinació d'aparença, comportament i funcionalitat– de manera "coherent, automàtica i homologable" per a diferents classes de documents. Gràcies a la participació de la Universitat de Leeds, CAMILEON va poder incorporar els resultats del projecte CEDARS.

Com es pot veure, les universitats (o, més ben dit, les seves biblioteques) van tenir un paper determinant en la posada en marxa de les primeres activitats: a la Gran Bretanya, mitjançant el JISC, i als Estats Units, mitjançant els esforços, en part, d'una de les seves associacions més emblemàtiques, l'RLG.

2.2. Iniciatives de biblioteques nacionals

Durant l'última meitat de la dècada dels noranta es van dur a terme algunes iniciatives destacables i influents; les tres de més ressonància internacional van ser les posades en marxa per les biblioteques nacionals d'Austràlia, Holanda i Suècia.

Austràlia

Mentre els britànics i els nord-americans iniciaven els seus primers projectes, a l'hemisferi sud la National Library of Australia (NLA) també es dedicava al tema de la preservació digital. A més dels avenços que va aconseguir, també tenen mèrit els seus esforços per promocionar la col·laboració entre institucions australianes i estrangeres. La seva contribució al coneixement de la problemàtica va ser gran i va incloure diversos aspectes. A continuació s'ofereixen algunes de les seves realitzacions més destacables: una selecció que no pretén ser exhaustiva, sinó donar una idea de la varietat d'activitats dutes a terme.

- **PADI** (*preserving access to digital information*): aquesta passarel·la, considerada una de les fonts d'informació més exhaustives sobre la preservació digital, dóna accés a recursos de tota classe, organitzats per temes (per exemple, formats i suports, captura de recursos web, propietat intel·lectual) i per tipus de recurs (per exemple, normes, vocabularis, llistes de discussió).
- **PANDORA**: es tracta d'un dels primers arxius de pàgines web. Va ser iniciat el 1996 per l'NLA i actualment hi col·laboren nou entitats australianes (biblioteques, arxius i museus) amb responsabilitats en la conservació del material cultural.
- **PANDAS** (PANDORA Digital Archiving System): tal com suggereix el seu nom, és el programari d'arxiu desenvolupat i utilitzat per PANDORA. La primera versió data del 2001 i actualment està en fase d'avaluació la tercera versió. L'NLA ofereix aquest sistema gratuïtament a altres institucions com, per exemple, el UK Web Archive, un projecte en què participen les tres biblioteques nacionals britàniques: la British Library i les d'Escòcia i Gal·les.
- *Directrius per a la preservació del patrimoni digital*: és un document oficial de la UNESCO, encarregat a l'NLA a causa dels seus coneixements i aportacions al camp de la preservació; es va publicar el 2003.

Holanda

La Koninklijke Bibliotheek (KB) ha possibilitat un avenç considerable dels coneixements del disseny, implantació i gestió de dipòsits digitals.

Les seves activitats han permès desenvolupar els seus propis serveis i han servit de model per a altres institucions de diferents països.

- **NEDLIB:** una de les primeres activitats de la KB va ser coordinar el projecte europeu NEDLIB. Aquest projecte, dut a terme durant el període 1998-2000, va tenir com a objectiu dissenyar una infraestructura que servís per a una futura biblioteca de dipòsit europea, que garantís "que les publicacions electròniques del present puguin ser utilitzades ara i en el futur". Entre les aportacions innovadores de NEDLIB hi ha el desenvolupament del prototip d'emulador, proposat per Jeff Rothenberg en el seu estudi *Avoiding technological quicksand*, i posat a prova durant el projecte CAMILEON. Altres biblioteques nacionals que també van participar en el projecte van ser les d'Alemanya, França, Noruega, Portugal i Suïssa.
- **E-Depot i Digital Information Archiving System (DIAS):** el sistema prototip desenvolupat en NEDLIB va conduir posteriorment al disseny i desenvolupament del dipòsit digital actual de la KB, E-Depot, que funciona sobre el sistema informàtic DIAS, encarregat per la KB a IBM. Aquest sistema gestiona recursos digitals, molt abundants i variats, de la preservació dels quals s'ha responsabilitzat la KB. Entre aquests hi ha les revistes científiques d'editorials internacionals tan prestigioses com Elsevier, Oxford University Press i Taylor & Francis, entre d'altres.

Suècia

El 1996 la Kungliga Biblioteket, igual que l'NLA, va començar a capturar les pàgines web d'interès nacional amb el projecte **Kulturarw3**. El seu objectiu va ser la captura *exhaustiva* de determinades classes de recursos del domini web suec, incloent-hi els diaris i les revistes digitals. Per a això van utilitzar des de molt aviat un programari que rastrejava periòdicament la xarxa per identificar i capturar les pàgines segons les pautes establertes. També va ser una de les pioneres a donar a conèixer alguns dels problemes legals que es presentarien a la preservació digital. La falta d'una legislació nacional que emparés la captura de pàgines web sense el consentiment dels titulars va posar en dubte la legalitat de les activitats. Finalment, un reial decret va autoritzar la Biblioteca Nacional a continuar capturant recursos mitjançant el seu robot i també a oferir el lliure accés al material –però només des de la biblioteca– per part dels lectors.

El projecte PADICAT (Patrimoni Digital de Catalunya), de la Biblioteca de Catalunya, iniciat el 2005, té com a finalitat capturar pàgines web considerades d'interès patrimonial, per a la seva preservació i accés permanent. El disseny i implantació de PADICAT han seguit les pautes i "millors pràctiques" identificades per les biblioteques nacionals australiana, sueca i danesa.

Lectura recomanada

Per a més informació sobre aquest projecte, vegeu **Ciro Lluca Fonollosa** (2006, novembre-desembre). "Archivando la Web, el proyecto Padicat (Patrimonio Digital de Cataluña)". *El Profesional de la Información* (vol. 15, núm. 6, pàg. 473-478). <http://eprints.rclis.org/archive/00007767/01/epi_padicat.pdf>

Els paràgrafs anteriors s'han limitat a esmentar només les principals institucions que, durant l'última dècada, han estat els referents per a altres institucions.

A més d'aquestes, moltes d'altres també han fet esforços valuosos en el camp de la preservació digital, algunes de les quals apareixeran esmentades en els mòduls successius.

A partir del 2000, es multipliquen les iniciatives i projectes, molts d'ells centrats en aspectes tècnics o administratius dins d'una determinada institució o relacionats amb un determinat tipus de fons digital. Alguns d'ells s'esmenten al llarg d'aquest text i en la bibliografia, però sense la pretensió de presentar una relació exhaustiva.

Continguts

Mòdul didàctic 1

Nou paradigma en preservació: la preservació digital

Alice Keefer Riva

1. Conceptes bàsics
2. Principals diferències entre la preservació tradicional i la preservació de recursos digitals
3. Preservació digital i preservació mitjançant la digitalització
4. Deteriorament col·lectiu i obsolescència tecnològica: antecedents

Mòdul didàctic 2

La preservació digital des de la perspectiva tècnica

Alice Keefer Riva

1. La informació digital: els seus elements i les seves amenaces
2. Les estratègies de la preservació
3. Els dipòsits digitals i la gestió del material custodiat

Mòdul didàctic 3

La preservació digital des de la perspectiva dels processos bibliotecaris

Alice Keefer Riva

1. L'adquisició de recursos digitals a les biblioteques: primer pas cap a la preservació
2. Les metadades
3. La gestió de la preservació

Mòdul didàctic 4

La preservació digital des de la perspectiva legal

Núria Gallart Marsillas

1. Dret d'autor: el marc legal
2. Les llicències comercials
3. Responsabilitats de la preservació
4. Recursos digitals que cal preservar

Mòdul didàctic 5

La preservació digital segons la tipologia

Alice Keefer Riva

1. Les dades científiques
2. Material en web
3. Les noves formes de comunicació en la Xarxa
4. Els objectes d'aprenentatge digitals

Bibliografia

Beagrie, Neil (2001). "Preserving UK digital library collections". *Program* (vol. 3, núm. 35, pàg. 217-226).

Cedars guide to digital collection management (2002). <<http://www.leeds.ac.uk/cedars/guideto/collmanagement/>>

Cedars guide to digital preservation strategies (2002). <<http://www.leeds.ac.uk/cedars/guideto/dpstrategies>>

Granger, Stewart; Russell, Kelly; Weinberger, Ellis (2002). "Cost elements of digital preservation". Ver. 4.
<<http://www.leeds.ac.uk/cedars/colman/costElementsOfDP.doc>>

Conway, Paul (1996). *Preservation in the digital world*. Commission on Preservation and Access. <<http://www.clir.org/pubs/reports/conway2/>>

Hedstrom, Margaret; Montgomery, Sheon. (1998). *Digital preservation needs and requirements in RLG member institutions*. Mountain View, Califòrnia: Research Library Group.
<<http://www.rlg.org/preserv/digpres.pdf>>

Lettenstrom, Frans (2000). "El Kulturarw³ Heritage Project: objetivos, desarrollo y planes futuros." A: *Biblioteques digitals i dipòsits nacionals de recursos digitals*. Barcelona: Universitat de Barcelona ("Quaderns de Treball", 15).

Long term preservation of electronic materials. Report of a JISC/British Library Workshop, Warwick, 27-28 November 1995 (1996). Londres: British Library.
<<http://www.ukoln.ac.uk/services/papers/bl/rdr6238/paper.html>>

Preserving digital information (1996). Mountain View, Califòrnia: Research Library Group.
<<http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>>

Rothenberg, Jeff (1999). *Avoiding technological quicksand: finding a viable technical foundation for digital preservation*. Washington: CLIR.
<<http://www.clir.org/PUBS/reports/rothenberg/contents.html>>

Rothenberg, Jeff (1995, març). "¿Son perdurables los documentos digitales?". *Investigación y Ciencia* (pàg. 8-13).