



**Secció oberta**

---

III Col·loqui Aula d'Història del Periodisme Diari de Barcelona:

Digitalització d'hemeroteques

*per Lluís Codina*

Professor titular de la Universitat Pompeu Fabra

Dins el procés general que podríem anomenar de digitalització de la cultura, la digitalització d'hemeroteques és, sens dubte, una de les dimensions més atractives que presenta aquest procés, si més no per als investigadors i professionals de la comunicació en particular i de les ciències socials en general.

Amb aquesta idea, l'Aula d'Història del Periodisme Diari de Barcelona (Estudis de Periodisme de la Universitat Pompeu Fabra i Arxiu Històric de la Ciutat de l'Ajuntament de Barcelona) va organitzar el 29 de maig d'enguany el seu III Col·loqui anual al voltant d'aquest tema, comptant amb les aportacions d'experts de la premsa, la documentació i els arxius i les col·laboracions de l'Associació d'Arxivers de Catalunya, el Col·legi de Bibliotecaris i Documentalistes, la Societat Catalana de Comunicació (IEC) i l'Observatori de la Comunicació Científica (UPF).

Els drets d'autor, els alts costos de la conversió del paper a *bits* i la preservació varen ser tres dels problemes principals abordats, tant per part dels ponents com per part del públic assistent, el qual va intervenir repetidament després de les exposicions de cada ponent.

## La propietat intel·lectual

Tot i que no es coneix el nombre amb exactitud, sí que se sap que hi ha moltes col·leccions digitalitzades de diaris i revistes, però la major part d'elles no es poden donar a conèixer al públic en general, ni encara menys es poden distribuir a través d'Internet. El motiu són les qüestions de propietat intel·lectual i de drets d'autor, aspectes que, en general, no estan ben resolts en el món digital.

En concret, d'una part, les empreses editores de les capçaleres tenen, lògicament, les seves reserves davant del fet que terceres empreses puguin obtenir qualsevol mena de benefici d'un producte la propietat del qual els correspon. Però, amb tot, no és aquest l'únic obstacle, sinó que també hi ha la qüestió dels drets dels autors individuals dels articles, que s'haurien de negociar un a un. Per tant, posar a l'abast una col·lecció digitalitzada de diaris o revistes implicaria entrar en un complicat laberint legal que, de moment, les empreses o organismes intenten evitar mantenint l'accés a les seves col·leccions a nivell intern o a nivell privat.

Les col·leccions antigues no tenen aquest problema, ja que la propietat intel·lectual desapareix en general 60 anys després de la mort de l'autor. En canvi, tenen el problema, força complicat tecnològicament, de la conversió del seu suport analògic (paper) a un suport digital (*bits*). Ferran Agelet, responsable de projectes de l'Arxiu de Barcelona, va presentar un informe sobre l'estat d'aquest problema.

## Lletres o imatges?

Conceptualment, el problema no sembla complicat: hi ha tecnologies que poden crear una reproducció facsímil dels documents en paper i convertir-la en una imatge digital (mapa de *bits*), de manera que els diaris queden convertits en *bits* i poden emmagatzemar-se en un suport informàtic i veure's en el monitor de l'ordinador; fins i tot, amb la tecnologia adequada, s'hi poden veure amb força fidelitat.

El problema és que, primer, la manipulació dels originals és complicada i consumeix una gran quantitat de temps, cosa que sol ser una barriera econòmica important. En segon lloc, amb aquest procediment, les col·leccions no són «cercables», és a dir, no són accessibles mitjançant recuperació d'informació. En tractar-se d'imatges, les col·leccions d'articles només es podrien consultar per desplaçaments seqüencials. Fins i tot, encara que es poguessin organitzar per capçaleres i per anys, al final caldria realitzar desplaçaments seqüencials, tal vegada entre milers de documents, per a trobar la informació.

La millor manera de fer una col·lecció de facsímils «cercable» consisteix a associar a cada document una fitxa bibliogràfica amb la referència del document i un conjunt de descriptors o paraules clau, però aleshores topem de nou amb un problema econòmic, ja que redactar aquestes fitxes és una feina que, a més de requerir personal ben entrenat, consumeix una gran quantitat de temps.

Hi ha més solucions, com la utilització de programes informàtics anomenats OCR (*Optical Character Recognition*) que converteixen les imatges facsímils dels documents en arxius de text. D'aquesta manera, és el propi text dels documents el que podria servir a efectes de recuperació d'informació. Ara bé, aquesta solució tampoc no és senzilla. Requereix igualment una considerable quantitat de temps a causa de la manipulació de la informació i de la correcció dels errors que cometem tots els sistemes d'OCR.

En aquest sentit, Marta Enrech, de la Unitat Documental de la Universitat Oberta de Catalunya, va explicar les bases de funcionament d'un suggestiu i interessant projecte de la UE anomenat LAURIN (Library and Archives Collecting Newspaper Clipping) en el qual, entre altres països europeus (7 en total), participa Espanya a través de la UOC.

LAURIN és de fet, tres projectes en un. D'una part, consisteix en un sistema informàtic integrat de *hardware* i *software* que digitalitza originals de diaris i els converteix també a text (amb un OCR), de manera que la base de dades que genera conté, al mateix temps, el facsímil (imatge) i el text complet de cada document.

En segon lloc, persegueix la creació d'una xarxa d'institucions, de ti-

pus arxivístic o museístic, que, de manera cooperativa, creïn bases de dades de premsa utilitzant el sistema LAURIN.

Per últim, el projecte consisteix en la definició d'estàndards aplicats a la creació de bases de dades de premsa, per exemple, sistemes d'indexació i de classificació, de codificació de les dades, etc.

## **Preservació i accés**

Anna Planet, gerent de la Biblioteca de Catalunya, va exposar els problemes relacionats amb la preservació de les col·leccions digitalitzades i l'accés a aquestes, i va explicar les línies mestres de la política de la Biblioteca de Catalunya en aquest sentit.

En primer lloc, fa ver una crida a la col·laboració entre totes les institucions que, d'una manera o altra, puguin estar involucrades en la digitalització de col·leccions de diaris o revistes.

La informació digital és summament fràgil. A més a més, la seva pròpia extraordinària abundància en dificulta la localització. Es requereixen normes clares i ben definides per al tractament documental i bibliogràfic de les col·leccions per tal que els documents que en formin part quedin ben identificats i que els ciutadans i els estudiosos puguin accedir-hi.

En aquest sentit, Planet va reclamar una política nacional de digitalització i la necessitat, entre altres coses, de crear un catàleg col·lectiu de fons digitalitzats. De moment, la BC està realitzant diversos estudis. Un d'aquests estudis es va realitzar mitjançant un qüestionari sobre la microfilmació i la digitalització de documents, que es va enviar a 261 centres de Catalunya. En aquest estudi es va posar en relleu que un 10 % dels centres tenien plans de digitalització de les seves col·leccions.

La Biblioteca de Catalunya està elaborant estudis sobre els criteris de selecció de documents a digitalitzar, és a dir, els criteris per a establir les prioritats i les diverses normatives involucrades en la digitalització, incloent-hi l'emmagatzematge, la transmissió de fitxers i la preservació. Es proposen donar suport en el futur a les biblioteques de Catalunya en temes de digitalització, assessorant en assumptes tecnològics, econòmics i legals.

## **Els centres de documentació dels diaris**

En el col·loqui varen intervenir dos representats dels centres de documentació de mitjans de comunicació: Carles Salmurri, per *La Vanguardia*, i Montserrat Rius, per *l'Avui*.

Va quedar de manifest que el panorama dels centres de documenta-

ció de premsa en el nostre país no és precisament exemplar pel que fa al grau de digitalització de les seves col·leccions històriques i a la seva posada a disposició del públic.

Tothom que hagi necessitat mai consultar un centre de documentació de premsa de qualsevol dels diaris esmentats podrà acreditar la seva eficiència general i la gran qualitat del seus serveis. Ara bé, això ho fan els diaris no gràcies als seus recursos informàtics, sinó malgrat la seva falta actual.

En general, el problema principal està relacionat amb dos dels aspectes mencionats al començament: en el cas de les col·leccions antigues no hi ha problemes de propietat intel·lectual, sinó de costos de digitalització. En canvi, la producció dels últims cinc o deu anys no té problemes de digitalització (ja està digitalitzada des de la seva creació), sinó de drets d'autor.

Tant el representant de *La Vanguardia* com el de *l'Avui* varen explicar que, afortunadament, estan en ple procés no sols de digitalització de les seves col·leccions històriques, sinó també d'estudi de la millor forma de facilitar-hi l'accés al públic (però no necessàriament gratuït, ni necessàriament a través d'Internet).

Lluís Codina, professor de documentació en els mitjans de la Universitat Pompeu Fabra, es va referir per la seva part a la necessitat de tenir una visió no només informàtica o documental sobre la qüestió, sinó també de tipus arxivístic o museístic per tal de garantir que les col·leccions digitalitzades ara siguin accessibles a les futures generacions; cosa que, de moment, no està tan clara si considerem que, a efectes reals, avui no podem ja recuperar documents enregistrats en els disquets de format 5" 1/4 que eren d'ús habitual fa poc més d'uns cinc anys i que, en canvi, encara podem llegir incunables de fa 500 anys.

També es va referir a la necessitat urgent de trobar fórmules jurídiques per al món digital, que defensin tant els drets de propietat intel·lectual –sense els quals no hi hauria un autèntic impuls a la creativitat– com els drets dels ciutadans a l'accés a la cultura, atès que es tracta d'una necessitat estratègica de les nacions per al seu avenç.

La presentació del Col·loqui i la seva ubicació dins les activitats de l'Aula d'Història del Periodisme Diari de Barcelona va anar a càrrec de Jaume Guillamet, catedràtic de periodisme, qui va destacar la importància de les col·leccions hemerogràfiques de premsa per a documentar la història social.

La taula rodona va estar moderada per Ramon Alberch, arxiver en cap de l'Ajuntament de Barcelona, qui va exposar alguns dels problemes habituals dels arxius històrics en la seva funció de suport fonamental en les activitats de recerca i docència en ciències socials i humanitats.

## *Algunes webs relacionades amb la digitalització d'hemeroteques*

### *Projecte LAURIN*

Projecte europeu de digitalització de diaris

<http://laurin.uibk.ac.at/>

### *L'Hemeroteca Electrònica (Hemeroteca de la UAB)*

Directorí internacional de publicacions periòdiques accessibles a través d'Internet

<http://www.bib.uab.es/hemerot/internet.htm>

### *Arxiu Municipal de Barcelona*

Informació i accés a l'Arxiu de l'Ajuntament de Barcelona, als seus projectes i publicacions

<http://www.bcn.es/arxiu/>

### *Biblioteca de Catalunya*

Informació i accés a la Biblioteca de Catalunya i al seu catàleg i directorí d'enllaços d'Internet de premsa (entre altres)

<http://www.gencat.es/bc/>

### *RLG DigiNews*

Publicació periòdica sobre col·leccions i arxius digitals

<http://www.rlg.org/preserv/diginews/index.html>

### *DLib Forum*

Publicació periòdica sobre biblioteques digitals

<http://www.dlib.org/>

### *Internet Library of Early Journals*

Col·leccions històriques del periodisme britànic

<http://www.bodley.ox.ac.uk/ilej>