

El laberinto de las hemerotecas digitales

Rafael Ibáñez / Bibliotecarios

Por motivos que no hacen al caso, durante los últimos meses me he visto embarcado en un proyecto para el que se requería la búsqueda sistemática de una información muy concreta y detallada en la prensa española de comienzos del pasado siglo y los últimos años del precedente. Cuando hace ya demasiados años debía realizar tareas similares, eso suponía la permanencia durante interminables jornadas en diferentes hemerotecas, quemándome las pestañas ante ingentes hojas de papel y rogando mayor diligencia al personal que debía proporcionarme los volúmenes (¡lo de la reproducción ya era para nota!).

Ha llovido mucho desde entonces y, por supuesto, Internet ha facilitado enormemente las cosas en este ámbito, como en otros muchos. Sin embargo, aun así, queda mucho por hacer (eso, o que con la edad voy incrementando mi nivel de exigencia, que también podría ser). El caso es que con la creación de las hemerotecas digitales podemos ahorrarnos muchos desplazamientos y estancias en ciudades que no son la propia, y los trabajadores de las instituciones que custodian los fondos ya no se ven impelidos a recorrer una y otra vez los depósitos cargados con esos pesados y delicados mamotretos. Pero lo cierto es que la fragmentación de los servicios digitales y la disparidad con la que atienden la normativa relativa a digitalización hacen que la tarea de los investigadores resulte ardua en exceso. Aunque, ciertamente, sentado ante mi ordenador, rodeado de objetos familiares y vistiendo sin etiqueta alguna, he podido afrontar la tarea prevista con mayor comodidad y rapidez que antaño —y con un considerable ahorro también, por supuesto—, lo cierto es que me he visto obligado a visitar diferentes hemerotecas digitales españolas.

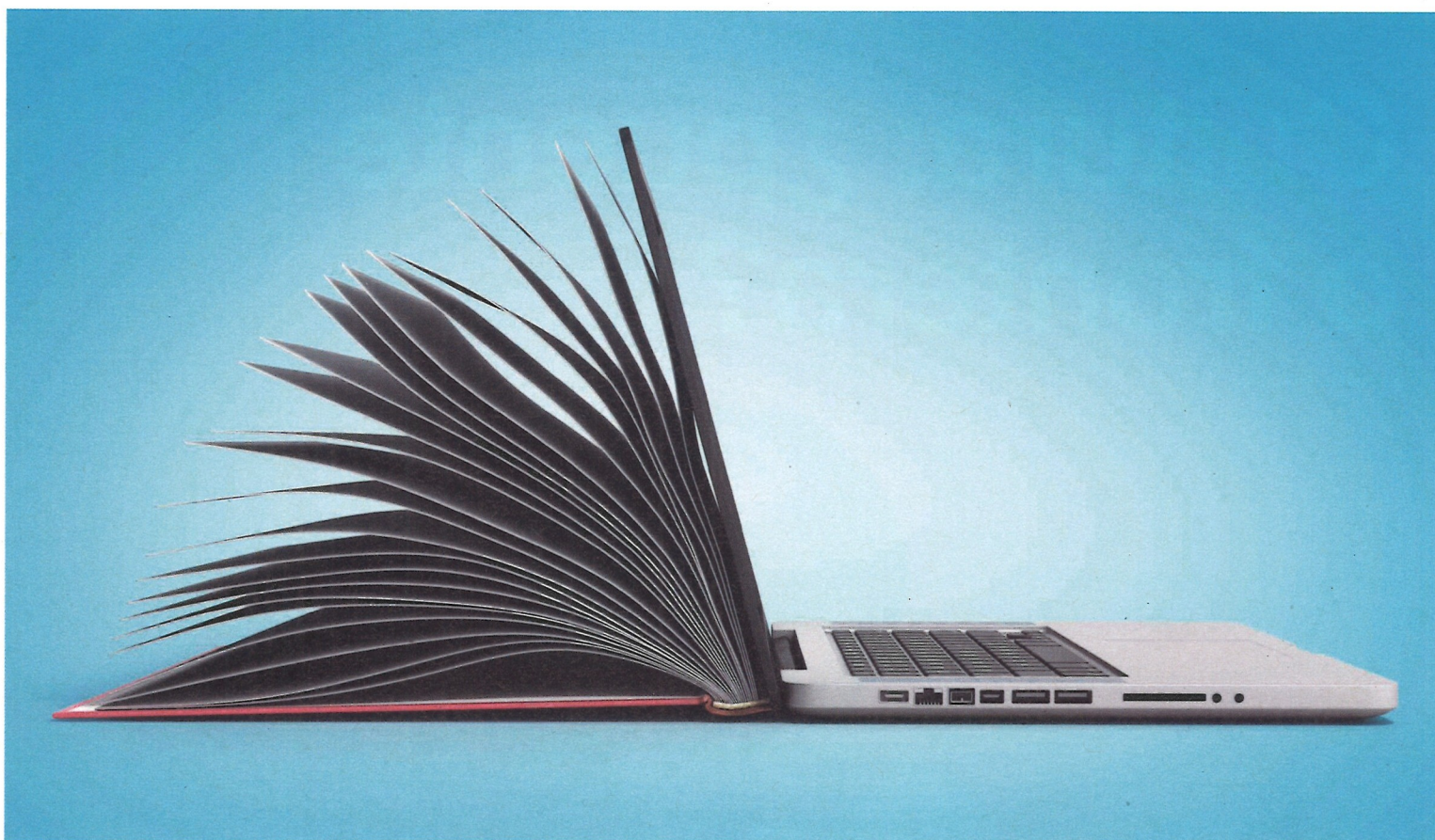
Tal vez lo más sangrante del caso haya sido comprobar cómo la Administración Central del Estado mantiene dos proyectos paralelos e independientes, ambos de una gran riqueza documental, pero cuya potencia —sin duda alguna— se vería exponencialmente incrementada si al menos el portal de acceso y los motores de búsqueda fuesen compartidos. Una de estas hemerotecas es la Biblioteca Virtual de Prensa Histórica, gestionada por la Subdirección General de Coordinación Bibliotecaria del Ministerio de Educación, Cultura y Deporte. Fruto de la cooperación del Ministerio con las Comunidades Autónomas, esta hemeroteca digital ofrecía, a fecha de hace un año, 7.541.157 de páginas de 2.369 cabeceras y 4.070 artículos que

proviene de colecciones de 97 instituciones (bibliotecas, archivos y museos, pertenecientes a comunidades autónomas, ministerios, ayuntamientos, ateneos, fundaciones, universidades, e instituciones privadas). Estas páginas están disponibles para su descarga en calidad media en formato JPG y PDF. La web ofrece unos formularios de manejo sencillo, que incluyen la búsqueda en el texto completo gracias a la aplicación del formato ALTO (Analyzed Layout and Text Object) para el proceso de reconocimiento óptico de caracteres.

Por su parte, la Biblioteca Nacional de España mantiene su propia Hemeroteca Digital, que forma parte del proyecto de la Biblioteca Digital Hispánica para la consulta y difusión pública a través de Internet del Patrimonio Bibliográfico Español conservado en la mencionada biblioteca. Su colección se ha ido ampliando progresivamente desde su creación en 2007, contando en la actualidad con 2.077 títulos y más de 56 millones de páginas de prensa digitalizadas. En este caso el formato de las publicaciones es únicamente PDF, aunque igualmente dotado con OCR, lo que permite buscar cualquier tema que se desee en el texto de la publicación. Pero me preguntó, ¿no tendría más sentido que ambos proyectos facilitasen la labor de los usuarios, unificando formularios y formatos, algo que sospecho no resultaría en exceso complicado?

Pero la locura no ha hecho más que comenzar. Con 374 de publicaciones cerradas, representativas de la cultura y la sociedad catalana, el Arxiu de Revistes Catalanes Antigueses es el otro gran depósito hemerográfico digital de acceso abierto existente actualmente en España, impulsado por la Biblioteca de Catalunya con el apoyo del Consorci de Serveis Universitaris de Catalunya (CSUC). Como en los casos anteriores, permite la búsqueda de términos en el contenido de las 2.402.710 páginas digitalizadas, pero en este caso la localización de los resultados es algo más ardua, entre otras cosas porque los documentos se han compilado en conjuntos dispares y la descarga de las páginas que interesen exige de nuevo su selección a través de un nuevo formulario.

Ya con estas tres referencias, cabe plantearse si la cooperación no podría avanzar más allá de la localización y digitalización de prensa original hacia la difusión y el acceso. Pero esto, claro está, no es todo: mientras la Generalitat de Catalunya cuenta con su propio buscador



a texto completo de prensa digitalizada en los archivos comarcales del Sistema d'Arxiu de Catalunya, a través del repositorio Trencadis la Diputació de Barcelona permite el acceso a los fondos hemerográficos de la Xarxa de Biblioteques Municipals y los Ayuntamientos de Gerona, Lérida y Tarragona hacen lo propio con los conservados en archivos municipales.

Incluso pueden mencionarse los casos de algunos convenios en la línea señalada en el párrafo anterior que por su escasa ambición provocan resultados que tal vez podrían calificarse como absurdos. Así, la Xarxa Vives d'universitats, de la que forman parte 22 universidades de la comunidad de cultura catalana en su más amplio sentido, distribuidas en la región Pirenaico-Mediterránea y pertenecientes a cuatro estados distintos (Andorra, España, Francia e Italia), cuenta con su hemeroteca digital. Con apenas 38 títulos, se encuentra integrada en la llamada Biblioteca Virtual Joan Lluís Vives, fruto de un convenio con la Fundación Biblioteca Virtual Miguel de Cervantes, que a su vez mantiene su propia hemeroteca digital, con algo menos de 300 títulos, entre los que se encuentran los aportados por la antedicha Xarxa.

Con sólo estas pinceladas resulta muy fácil hacerse a la idea del panorama de las hemerotecas digitales en España (por lo que he podido ver, no muy diferente del francés, por ejemplo). Son varias las comunidades autónomas que cuentan con su hemeroteca digital, ya sea mantenida por la propia Administración Autonómica (como es el caso de la hemeroteca en la Biblioteca Virtual Andalucía, de la hemeroteca en la Biblioteca Virtual del Principado de Asturias o la hemeroteca en Galicia), en colaboración con otras administraciones y entidades (así ocurre con la Euskal Prentsaren Lanak) o por instituciones culturales y de estudios locales (como sucede con la sección de prensa de la Biblioteca Virtual de Castilla-La Mancha o Jable, el archivo canario de

prensa digital). Obviamente, también muchos Ayuntamientos han puesto en marcha su hemeroteca digital, ya sea partiendo de un prestigioso servicio hemerográfico como el de Madrid, que ha integrado sus fondos digitalizados en la biblioteca digital Memoria de Madrid (utilizada la opción 'Hemeroteca Municipal' del menú desplegable Instituciones para seleccionar los fondos de prensa); de las colecciones conservadas en sus archivos, (como han hecho los de Toledo, Huelva, Cartagena y Murcia) o en sus bibliotecas, que es el caso de San Sebastián.

También algunas cabeceras actuales pero que cuentan ya con una larga trayectoria vital ofrecen gratuitamente el acceso a la digitalización de su colección. En algunos casos, como los de ABC o La Vanguardia, permiten búsquedas en el contenido de sus objetos digitales, devolviendo los resultados en formato PDF. En cambio, otros medios como El País solamente devuelven los contenidos (no las páginas) en formato HTML. Mientras tanto, algunas cabeceras desaparecidas cuentan en la web con su versión digital gracias a proyectos editoriales autónomos como el de Ediciones Pléyades, que de esta forma ha recuperado los últimos veinte años de Triunfo y la revista Tiempo de historia.

Tal pluralidad de iniciativas y proyectos se traduce no sólo en una incomodidad para el investigador y el curioso, que ha de rastrear numerosas plataformas y batallar con diferentes diseños, sino también en una enorme disparidad de formatos que no siempre permitirán la confluencia de esfuerzos. Si en algunos de los casos citados los proyectos se ajustan a normas que permiten la recolección de sus objetos por Hispana, en otros este propósito resulta imposible. Más pronto que tarde debería imponerse la cordura para que la localización y el acceso a la riqueza informativa de nuestra prensa histórica no requiera un desmedido esfuerzo, impropio de la era tecnológica en que vivimos.