PRESERVACIÓN DIGITAL Y BIBLIOTECAS EN EUROPA

Carmen Díez Carrera[1]

Recibido: 31/06/2013 - Aprobado: 10/09/2013

RESUMEN
En el artículo se reflexiona acerca del valor de la preservación en la gestión del patrimonio digital, a través del estudio de caso de tres bibliotecas nacionales europeas: la Biblioteca Nacional de España, la *BibliothèqueNationale de France* y la *British Library*. Se analizan y comparan la visibilidad y alcance de las políticas, el grado de desarrollo e implantación, las técnicas de preservación incluyendo los esquemas o estándares empleados y los modelos de metadatos, los planes de prevención de riesgos, la participación en proyectos nacionales e internacionales, la protección de la propiedad intelectual y de los derechos de autor, y el cálculo de costes de la preservación a medio y largo plazo.

Palabras clave: Bibliotecas digitales. Preservación digital. Bibliotecas Nacionales. Costes de la preservación. Políticas de preservación digital.

ABSTRACT
This article aims to reflect the strategic importance of preservation in digital asset management, through a case study of three european national libraries representing the European Union average: Spain, France and United Kingdom. We analyze and compare the visibility and scope of policies, the degree of development and implementation, including preservation techniques and materials, used schemesand metadata models, risk prevention plans, participation in national and international projects, protection of intellectual property and copyrights, and costing of the preservation medium and long term.

Keywords: Digital libraries. Digital preservation. National libraries. Costs of preservation. Digital preservation policies.

1. Profesora Titular de Biblioteconomía y Documentación. Universidad Carlos III de Madrid c/ Madrid, 126 - 28903 Getafe (España) carmen.diez@uc3m.es

QUÉ ENTENDEMOS POR PRESERVACIÓN DIGITAL

El origen de este texto se dio hace unos pocos años con motivo de la preparación de contenidos prácticos docentes, sobre los que posteriormente he seguido reflexionando para concluir con este trabajo como resumen. Se trata de un estudio de caso, basado en el análisis de las webs de las bibliotecas seleccionadas, y en el que se busca respuesta a ciertas preguntas sobre si las bibliotecas nacionales europeas son conscientes de los retos de la preservación digital y, en caso afirmativo, cuáles son las políticas y las realizaciones al respecto. En consonancia con este carácter práctico, se ha evitado la erudición bibliográfica. Todos los enlaces que aparecen en el artículo han sido comprobados a 20 de junio de 2013.

Ante la rapidez con que se crean repositorios de objetos digitales, nacidos así o digitalizados, es necesario adaptar estándares para que los contenidos resulten útiles, fiables, auténticos y accesibles, para superar el reto de garantizar la accesibilidad y preservación del patrimonio cultural.

Aunque parezca que se trata solo de dos tipos de objetos, hay que distinguir al menos los siguientes:

- Publicaciones en línea o en red recibidas mediante correo electrónico o ftp (Protocolo de Transferencia de Ficheros).
- Publicaciones digitales en CD-ROM.
- Discos anexos de publicaciones impresas.
- Imágenes digitales.
- Datos de sistemas de información.
- Proyectos de digitalización que producen imágenes y productos interactivos.
- Sitios Web.
- Grabaciones sonoras.
- Correos electrónicos, discos, cintas y otros medios que forman parte de archivos personales y colecciones de manuscritos.

A diferencia de la colección en soportes tradicionales, cuya preservación se centra en garantizar la integridad material de sus componentes, la de los objetos digitales se centra en garantizar la accesibilidad a lo largo del tiempo. Accesibilidad y tiempo, dos aspectos que tienen connotaciones específicas y radicalmente distintas que en el mundo analógico.

Las *Directrices para la Preservación del Patrimonio Digital de la UNESCO* [http://unesdoc.unesco.org/images/0013/001300/130071s.pdf- definen la preservación digital incidiendo en el primer aspecto:

> *«La preservación digital consiste en los procesos destinados a garantizar la accesibilidad permanente de los objetos digitales. Para ello, es necesario encontrar las maneras de representar lo que se había presentado originalmente a los usuarios mediante un conjunto de equipos y programas informáticos que permiten procesar los datos. Para lograrlo, es necesario que la comprensión y la gestión de los objetos digitales se realice considerándolos desde cuatro puntos de vista: como fenómenos físicos, como codificaciones lógicas, como objetos conceptuales comprensibles para el ser humano y como conjuntos de elementos esenciales que deben ser preservados para ofrecer a los futuros usuarios lo esencial del objeto.»*

Planteada en términos de permanencia en el tiempo, la preservación está condicionada por la rápida caducidad de los elementos que hacen accesibles los documentos (*hard* y *soft*), lo que introduce una nueva dimensión al factor tiempo. Tradicionalmente éste ha sido considerado en términos de siglos como medida de los objetivos a largo plazo, incluso en condiciones adversas. En el entorno digital la incertidumbre hace imposible un planteamiento así, y el concepto de largo plazo se reduce a unas pocas décadas, y es acompasado por objetivos a corto y medio plazo, más adecuados a la naturaleza de los entornos digitales.

La preservación digital es esencialmente preservar el acceso en el tiempo. Esto hace virtualmente imposible segregar con nitidez los costes de la preservación digital, de los costes del acceso.

Los costes del acceso son significativos, porque tanto la tecnología, como las expectativas de los usuarios, avanzan a un ritmo muy rápido. Se espera que los costes de preservación digital sean mayores que los del papel, por lo siguientes factores:

1. La necesidad de gestionar activamente los cambios inevitables en la tecnología a intervalos regulares, y hacerlo en un plazo de tiempo potencialmente infinito.
2. La falta de normalización en los recursos mismos hace que sea difícil lograr economías de escala.
3. El tema aún sin resolver de los medios para representar de forma fiable y precisa los objetos digitales, demostrando que no han perdido información esencial después de los cambios tecnológicos.

4. Durante algún tiempo la preservación digital va a ser un coste adicional, añadido al de los costes de conservación de las colecciones tradicionales. Las bibliotecas que custodien ambos se enfrentan al reto más difícil, pues necesitarán equilibrar el reparto de los recursos entre dos exigencias tan distintas. Sobre todo teniendo en cuenta que estas instituciones acostumbradas a priorizar la conservación a largo plazo, van a tener que priorizar también el acceso a corto plazo.

Si bien existe una preocupación comprensible porque los costes de la preservación digital van a ser altos, es igualmente importante considerar que los costes y las implicaciones de la no conservación de los objetos digitales serán igualmente importantes. Los costes de recrear un objeto digital pueden ser mucho mayores que los de conservación; además de que puede que no haya siquiera posibilidad de hacerlo transcurrido cierto tiempo (http://www.dpconline.org/advice/preservationhandbook/digital-preservation/strategic-overview)

En este sentido las políticas de la biblioteca digital se deben centrar en dos grandes objetivos: la preservación a largo plazo de objetos digitales íntegros y accesibles. Objetivos que deben ser contemplados desde el momento de la creación del objeto digital, mediante el uso de los metadatos adecuados (p. ej. PREMIS) que garanticen su preservación y accesibilidad en el tiempo.

Las estrategias de preservación digital buscan resolver los problemas de la obsolescencia tecnológica en el tiempo, ya que si ésta se mueve en el corto plazo (cada vez son más cortos los periodos de vigencia tecnológica, en torno a los 3 años), aquéllas buscan soluciones en el medio y largo plazo (en términos de tecnología la conservación a medio plazo está sobre los 5 o 7 años y a partir de aquí se considera largo plazo; poco que ver con el mundo en papel).

Para hacernos una idea de la dimensión que todo esto supone baste con la constatación de que entre 2005 y 2011 se ha multiplicado por 8 la información creada digitalmente, mientras que la información accesible sólo se ha multiplicado por 4 (http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf).

En función de los medios y de las líneas estratégicas se emplean unas u otras técnicas, entre las que destacan:

5. Las copias de seguridad (*backup*) en soportes duraderos: técnica empleada sobre todo para los sistemas de gestión, con buenas cotas de calidad de imagen, como de durabilidad de los soportes. Una versión es el repicado distribuido, que emplea múltiples servidores remotos sobre los que se descargan las copias. En bibliotecas existe

la solución LOCKSS (*Lots of Copies KeepStuffSafe*) consultable en: http://lockss.stanford.edu/lockss/Home.

6. La renovación: genera una reproducción completa tanto del contenido como de los elementos formales de un documento, la cual puede ser considerada copia fidedigna del original.
7. La migración: es la transferencia periódica entre configuraciones o entre generaciones de tecnología, que permite conservar la información en un modo que no garantiza ser una réplica exacta del original. Reproduce el contenido del documento, pero cambia la configuración y el formato.
8. La emulación: reproduce el comportamiento y los resultados de un *hardware* o de un *software* desaparecidos para recuperar documentos y datos, tanto de aplicaciones, como de sistemas informáticos obsoletos. También es conocida como arqueología digital.
9. La conservación de la tecnología: conservar los sistemas (*hard* y *soft*) después de haber sido sustituidos. Es una alternativa cercana a un museo de la tecnología, costosa y poco viable, que representaría la plena disponibilidad con independencia del tiempo transcurrido.

LAS POLÍTICAS EN LAS BIBLIOTECAS NACIONALES DE ESPAÑA (BNE), FRANCIA (BNF) Y EL REINO UNIDO (BL)

Para contestar a la cuestión se han elegido tres centros representativos y analizado los aspectos más relevantes: la visibilidad y alcance de las políticas, el grado de desarrollo e implantación, las técnicas de preservación incluyendo los esquemas o estándares empleados y los modelos de metadatos, los planes de prevención de riesgos, la participación en proyectos nacionales e internacionales, la protección de la propiedad intelectual y de los derechos de autor, y el cálculo de costes de la preservación a medio y largo plazo.

Las bibliotecas española y francesa tienen en común la **visibilidad** de los recursos digitales con acceso directo a las herramientas: Biblioteca Digital Hispánica y Hemeroteca Digital, y *Gallica* respectivamente. En la inglesa y en todas para los demás recursos digitales se accede a la información navegando a partir de alguna de sus pestañas principales y siguiendo criterios diferentes. También es común a las tres que las políticas forman parte de la información corporativa, mientras que las prácticas y las cuestiones técnicas se alojan en otros sectores. A partir de aquí no existe un criterio para ubicar la información: la BNE la sitúa en la sección denominada "Catálogos" (http://www.bne.es/es/Catalogos/) bajo el epígrafe Biblioteca Digital Hispánica, incluyendo la Hemeroteca digital y el proyecto ENCLA-

VE para el acceso a las referencias y al editor de obras digitales sujetas a derechos de autor; mientras que el acceso a la iniciativa Archivo de la web española para la conservación del dominio punto es (.es) debe hacerse a través del buscador, pues carece de acceso navegable (http://www.bne.es/es/LaBNE/PreservacionDominioES/). La BNF la aloja en la sección destinada a los profesionales (http://www.bnf.fr/fr/professionnels/numerisation.html), separada de los puntos de acceso que son dos: *Gallica*, alojada en la sección de "*Colections et services*" (http://www.bnf.fr/fr/collections_et_services/bibliotheques_numeriques_gallica.html), y la Biblioteca digital de los niños, en "*Evenements et culture*" (http://enfants.bnf.fr/); la BL la ubica en el apartado institucional "*Aboutus*", dentro del apartado "*Strategy, policies and programes*" bajo lo epígrafes "*Collectioncare*" y "*The digital programme*" (http://www.bl.uk/aboutus/stratpolprog/digi/index.html), mientras que el acceso se hace a través de la pestaña "*Collections*" en el apartado "*Digital collections*" (http://www.bl.uk/#). Es muy llamativa la falta de criterios homologables para distribuir y acceder a la información en instituciones culturales que son básicamente equiparables.

El **alcance de las políticas** tiene en común unos mismos objetivos: la digitalización selectiva de la colección analógica de obras huérfanas sobre todo, los objetos digitales en un sentido muy amplio que incluye los derechos de propiedad intelectual, y las páginas web de sus respectivos dominios nacionales, además de participar en proyectos domésticos e internacionales. A partir de esta base las diferencias son notables en enfoque, dimensión y trayectoria. Mientras que la francesa y la británica se pusieron a la tarea a comienzos de la década de los 90, la española todavía en el Plan Estratégico 2006-2008 se estaba planteando dar los primeros pasos reducidos a digitalizar un centenar de obras y a dotarse de la infraestructura necesaria para alojar la futura biblioteca digital.

Hay un aspecto en el que las tres, y en general las demás bibliotecas, muestran que la gestión de lo digital tiene aun un aire de excepcionalidad, pues tras más de dos décadas su tratamiento está segregado del general de la colección, estableciendo con ello dos colecciones de hecho en función del soporte y del medio de registro. Es un síntoma de que el entorno digital sigue en fase introductoria, aun y así cabe preguntarse ¿qué justifica la segregación de contenidos desde el punto de vista del servicio?

Grado de desarrollo e implantación. La BNF posee más de 324.000 libros, cerca de 21.000 manuscritos, más de 22.000 cabeceras de prensa y 2,5 millones imágenes digitalizadas (http://gallica.bnf.fr/), la BL no ofrece cifras globales aunque en los últimos tres años ha realizado más de 16 millones de digitalizaciones, la BNE más de 54.000 obras digitalizadas (http://bdh.bne.es/bnesearch/) a las que habría que añadir los cerca de 4 millones de páginas de la Hemeroteca digital. Llama una vez más la atención la carencia de una política transparente de información, obtener cifras es

muy costoso y los resultados son parciales y en unidades de medida que no son comparables.

Además de la distancia que revelan las cifras, el alcance de las políticas presenta una diferencia equiparable, y mientras que la BNF y la BL poseen políticas robustas y con más de veinte años de experiencia, la BNE está bastante lejos y su Plan Estratégico 2012-2014 hace un análisis DAFO en el que incluye la ausencia de un plan gestor de digitalización entre las debilidades estructurales (http://www.bne.es/es/LaBNE/MemoriaPlan/PlanEstrategico/index.html).

La BL es con mucho la que posee una política de preservación digital de mayor alcance. Está integrada en la estrategia general de preservación, dentro de la cual posee un grupo de trabajo (*Digital PreservationTeam*) centrado en los problemas de la obsolescencia y en desarrollar soluciones que aseguren la longevidad de su patrimonio, entre ellas: análisis de las tecnologías, elaboración de planes (como el de *MS Office* desarrollado con Microsoft y el de prensa digital), planificación de riesgos, creación de un banco de pruebas y la implicación en proyectos nacionales e internacionales (http://www.bl.uk/aboutus/stratpolprog). Es la única que se ha planteado proporcionar acceso a las especificaciones de los formatos de archivo digitales, lo ha comenzado con Microsoft, pues será un aspecto crítico para el acceso futuro a la información cuando las aplicaciones actuales se hayan convertido en obsoletas.

También difieren en las **técnicas de preservación** y en la información que dan sobre el tema. La BNE habla del control de calidad como objetivo del próximo bienio (2012-2014) y del uso de metadatos para asegurar la accesibilidad, pero nada trasluce de la existencia de una planificación, ni de cuales sean las prácticas. La BNF ha puesto en marcha el sistema SPAR (*Systéme de Préservation et d'ArchivageRepartie*), un sistema de almacenamiento distribuido (*LOCKS* ya mencionado) que cumple la norma OAIS (ISO 14721:2012) y realiza numerosas copias de los objetos digitales, con procedimientos de vigilancia tecnológica para actualizar formatos, y que permite volver siempre al formato original ante cualquier contingencia (http://www.bnf.fr/fr/professionnels/conserver_spar/s.conserver_SPAR_presentation.html). La BL tiene el programa DOM (*Digital Object Management*), un sistema muy completo también de almacenamiento distribuido de copias por *backup* en las bibliotecas nacionales de Gales y Escocia, con un control de integridad para asegurar que no se pierden datos por deterioro o por catástrofe y seguimiento de versiones y accesibilidad (http://www.bl.uk/aboutus/stratpolprog).

Las tres bibliotecas coinciden en el uso de **estándares y esquemas** como Marc21, OAI_PMH, *DublinCore*, PREMIS y METS, que son el soporte para el desarrollo de herramientas y procesos de sus respectivos sistemas

SPAR y DOM. Por lo mismo coinciden en el USO de PREMIS como esquema de **metadatos para la preservación** digital basado en el modelo OAIS.

Fuera del uso de los estándares generalmente aceptados las diferencias vuelven a ponerse de manifiesto cuando analizamos la **existencia de planes de prevención de riesgos**, algo sobre lo que la BNE no proporciona información, posiblemente porque aún no hayan sido desarrollados, y que la BNF tampoco especifica aunque están implícitos en la filosofía del sistema SPAR. La BL viene implementando desde 2007 un exhaustivo programa en constante evolución, que comprende todos los formatos y todas las colecciones (http://www.bl.uk/aboutus/stratpolprog) y para ello sigue la norma australiana y neozelandesa AS/NZS 4360:2004 *Risk Management standard*.

La **participación en proyectos nacionales e internacionales** de preservación digital es otro aspecto en el que destaca la BL, miembro fundador de la DPC (*Digital PreservationCoalition*), del IIPD (*International Internet PreservationConsortium*) al que también pertenecen las otras bibliotecas. Entre los principales proyectos internacionales están: LIFE (*LifecycleInformationfor E-Literature*) que ha desarrollado un modelo de ciclo de vida digital y el modelo para el cálculo de costes de preservación, PLANETS (*Preservation and Long-term Access throughNETworkedServices*) para crear un marco sostenible de preservación digital a largo plazo, IMPACT (*IMProving Access to Text*) orientado a eliminar las barreras al acceso a los textos históricos digitalizados, PREMIS (*PREservation Metadata Implementation Strategies working group*), PRESERV2 sobre preservación de repositorios e interoperabilidad. La BNF participa en IMPACT y además en KEEP (*Keeping EmulationEnvironments Portable*) cuyo objetivo es constituir una plataforma de software para garantizar el acceso a largo plazo a los contenidos multimedia. Por la información que proporciona, la BNE se limita a participar en el proyecto IMPACT.

Otro aspecto en el que coinciden básicamente es el de la **protección de la propiedad intelectual y los derechos de autor**. La digitalización se centra en obras huérfanas y lo que se proporciona en otro caso es la ficha catalográfica con un resumen, un fragmento de la obra y acceso al editor y a su sistema de compra en línea. Donde difieren es en la visibilidad y en el desarrollo de las políticas, en la BNE cuesta encontrar información salvo que se sepa previamente que ENCLAVE es el proyecto desarrollado con los editores españoles para su aplicación en la Biblioteca Digital Hispánica. La BNF tiene abundante información bien identificada tanto en la sección profesional como en la propia Gallica, y su política se resume a respetar la propiedad y los derechos digitales ofreciendo también un sistema consorciado con los editores, para que a través de la descripción catalográfica se acceda a las plataformas de compra. La diferencia está en

que la biblioteca puede digitalizar obras protegidas, a las que da acceso individualizado en sus propias instalaciones exclusivamente, lo mismo que para obras digitales. La BL ofrece abundante información, reflexiones y recomendaciones a los poderes públicos para actualizar la legislación sectorial, aunque por ahora sus realizaciones no van más allá de lo indicado por la presencia de intereses en conflicto. (http://pressandpolicy.bl.uk/content/default.aspx?NewsAreaID=17).

Por último, en ninguna de las tres se ofrecen datos del **cálculo de los costes de la preservación digital**, sino solamente de la digitalización y de las infraestructuras con diferente grado de detalle, y lo más llamativo es que formando parte la BL del proyecto LIFE que ha desarrollado un modelo para dicho cálculo, no lo ha aplicado a su propio caso.

A modo de conclusión sería recomendable que en el contexto de los grupos de coordinación y encuentro de las bibliotecas nacionales de la UE se buscara algún tipo de armonización en materia de preservación digital especialmente en criterios que guíen las políticas y la estructura de la información dada al público, especialmente empleando esquemas y unidades de medida normalizados que faciliten el análisis y la comparación. Asimismo es vital para la viabilidad de la preservación digital encarar un estudio a gran escala sobre los costes de la misma, teniendo en cuenta que va camino de hacerse insostenible en un entorno de recursos decrecientes. Se trata de poder planificar el futuro inmediato de acuerdo con un orden de prioridades respecto de lo que será el patrimonio digital que reciban las generaciones futuras. No olvidemos que la cuestión planteada se desenvuelve en un entorno que se desarrolla y cambia a gran velocidad, en el que las tecnologías se desfasan dejando poco o ningún espacio a la compatibilidad. Recordemos que aunque los soportes pueden durar un siglo o más, caso del CD-ROM, no hay sistema (*hard* y *soft*) utilizable durante la vigésima parte del tiempo.