

Patrimonio digital: a vista de pájaro

Belén ALTUNA ESTEIBAR*

La *Carta sobre la preservación del patrimonio digital*, de la UNESCO¹, define a este último como “recursos únicos que son fruto del saber o la expresión de los seres humanos. Comprende recursos de carácter cultural, educativo, científico o administrativo e información técnica, jurídica, médica y de otras clases, que se generan directamente en formato digital o se convierten a éste a partir de material analógico ya existente. Los productos de *origen digital no existen en otro formato que el electrónico*”.

Es precisamente de los proyectos que se ocupan de la recogida y conservación de objetos cuya existencia es exclusivamente electrónica de lo que va a tratar este artículo, no de una manera exhaustiva, sino una panorámica general y básica del estado de la cuestión.

Primeras iniciativas

Allá por la “prehistoria” de la www —el primer navegador *Mosaic* aparece en 1993—, ya empieza a haber no sólo inquietudes sino también experiencias sobre qué hacer, cómo hacer, para almacenar y conservar ese todavía incipiente tipo de documento, que en adelante y más genéricamente trataremos como “objeto digital”. No hay que olvidar que tan solo catorce años después, en 2007 hay más de 400 millones de dominios registrados en la web, según el Internet Systems Consortium².

En 1994, en Canadá aparecen los primeros proyectos de preservación de patrimonio digital de que se tiene noticia. Como decíamos más arriba, no se trata de hacer una revisión histórica exhaustiva pero sí hay que señalar que para el año 2000 Países Bajos (1995), Australia, Suecia (1996), Alemania, Finlandia, Islandia (1997), Dinamarca (1998), Austria, Nueva Zelanda (1999), Francia y Estados Unidos (2000) han puesto en marcha proyectos, más o menos parciales, más o menos acertados, más o menos sostenibles para intentar enfrentar una situación realmente desconocida anteriormente y cuya mayor parte de variables son incontroladas en el momento (incertidumbre sobre estándares, inadecuación de legislaciones regulatorias existentes, fragilidad tecnológica tanto desde el punto de vista de las aplicaciones

49

* Actualmente ocupa el puesto de Subdirección de la Biblioteca de la Universidad Pública de Navarra, belen.altuna@unavarra.es

1. *Carta sobre la preservación del patrimonio digital*. UNESCO, 2003. (Disponible en: http://portal.unesco.org/es/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html). [Consulta: 5 de octubre de 2007].

2. Disponible en: <http://www.isc.org/index.pl?ops/ds/>. [Consulta: 5 de octubre de 2007].

como de los protocolos..., simultáneamente, crecimientos exponenciales de objetos digitales que cada vez en mayor número son solamente digitales (no hay un CD, un DVD, un disquete o cualquier otro soporte físico que mantenga, al menos temporalmente, la información), son cada vez más dinámicos y, simultáneamente, más efímeros: nacen, crecen, se desarrollan y mueren en la www... sin dejar rastro.

A fecha de hoy, aunque en el ámbito de las tecnologías, en general, es muy arriesgado ser pionero, parece que el desarrollo y las tendencias tecnológicas les han dado la razón y se manifiesta que muchos de los países que fueron sensibles y se enfrentaron a la situación desde el principio, lideran proyectos que son canónicos desde el punto de vista del "qué y cómo hacer" en el área de *preservación de patrimonio digital*, y entre ellos el proyecto PANDORA australiano, el KOPAL alemán, el Kulturarw3 sueco, el NDEP holandés, entre otros.

Aunque del caso español hablaremos más adelante cabe señalar que, por el momento, sólo Catalunya ha comenzado a enfrentar la conservación de la producción digital y ello, recientemente, en el año 2005.

Implicación de organismos internacionales

Estas iniciativas que fueron surgiendo de modo disperso, van dejando oír su voz en organismos internacionales culturales como la UNESCO, en organismos supranacionales como la UE, organizaciones profesionales como la IFLA (International Federation of Library Associations) y empresariales como la IPA (International Publishers Association).

50

La UNESCO aprueba en 2003 la *Carta sobre la preservación del patrimonio digital*³, mencionada en la introducción de este artículo, en la que se considera que el "patrimonio digital se encuentra en peligro de desaparición, y que su preservación en beneficio de las generaciones actuales y futuras es una preocupación urgente en el mundo entero" e insta a que los gobiernos y los estados definan medidas y estrategias para proteger y preservar el patrimonio digital. En el artículo 9 se hace hincapié en que se debe preservar de modo que se propicie que en el futuro se disponga de un patrimonio que constituya "una representación de todos los pueblos, naciones, culturas e idiomas" y se pone de manifiesto, entre otras cosas, la necesidad de colaboración entre distintos agentes sectoriales relevantes, tanto tecnológicos, como de instituciones de la memoria, empresariales, etc.

Esta carta fue precedida, en marzo de 2003, por la publicación de las *Directrices para la preservación del patrimonio digital*⁴, documento elaborado por la Biblioteca Nacional de Australia para la División de la Sociedad de la Información de la UNESCO, en la que se deta-

3. *Carta sobre la preservación del patrimonio digital*. UNESCO, 2003. (Disponible en: http://portal.unesco.org/es/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html). [Consulta: 5 de octubre de 2007].

4. Biblioteca Nacional de Australia. *Directrices para la preservación del patrimonio digital*. UNESCO, 2003. (Disponible en: <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>). [Consulta: 5 de octubre de 2007].

llan a lo largo de más de un centenar de páginas, aspectos técnicos, prácticos y de gestión para poder llevar a cabo un proyecto de preservación de patrimonio digital con unas mínimas garantías de éxito.

Si bien el plan de acción eEurope 2002, que recibe el apoyo de los estados miembros en Portugal en el año 2000, incluye, entre otros, el objetivo de “*estimular los contenidos digitales europeos en las redes mundiales*”⁵, se refiere todavía más a *contenidos digitalizados que al ámbito de objetos exclusivamente digitales*, del que nos estamos aquí ocupando. De hecho, los *Principios de Lund*⁶ a los que profusamente se alude al hablar de cuestiones digitales se están refiriendo precisamente a la coordinación europea de programas de digitalización retomando el plan eEurope.

Es en el seno del Consejo de la Unión Europea en junio de 2002, cuando se aprueba una resolución cuyo título es *Conservar la memoria del mañana-Conservar los contenidos digitales para las generaciones futuras*⁷ en la que se constata la fragilidad del soporte y la rápida obsolescencia de las tecnologías en las que se produce cada vez una mayor parte de los “valores culturales e intelectuales de nuestra sociedad”. Se hace mención explícita a la necesidad de apoyar a los organismos de conservación (bibliotecas, archivos y museos) “en sus responsabilidades a la hora de recoger el contenido y mantenerlo accesible a lo largo del tiempo”. A pesar de esta manifestación contenida en la resolución mencionada, el énfasis a nivel europeo se diría que parece ponerse más en los proyectos de digitalización que en el patrimonio únicamente digital.

Ahora bien, en el mes de septiembre de 2005, la Comisión adopta la Comunicación⁸ *i2010: Bibliotecas Digitales*, que surge al calor de la iniciativa de varios países europeos sobre la creación de una biblioteca digital europea (TEL, *The European Library*) intentando contrarrestar el anuncio de Google de digitalización de 15 millones de libros conservados en cinco bibliotecas del ámbito anglosajón, en aquella fecha y que a día de hoy es un proyecto en el que además de siete bibliotecas americanas, se han incorporado bibliote-

51

5. *eEurope 2002. Una sociedad de la información para todos. Plan de acción preparado por el Consejo y la Comisión Europea para el Consejo Europeo de Feira, 19-20 junio 2000. Bruselas: Comisión de las Comunidades Europeas, 2000.* (Disponible en: http://europa.eu.int/information_society/europe/2002/action_plan/pdf/action_plan_es.pdf). [Consulta: 5 de octubre de 2007].

6. *eEurope 2002: Digitising content together. Los principios de Lund: conclusiones de la reunión de expertos, Lund, Suecia, 4 de abril de 2001.* (Disponible en: ftp://ftp.cordis.europa.eu/pub/ist/docs/digicult/lund_principles-es.pdf). [Consulta : 5 de octubre de 2007].

7. Consejo de la Unión Europea. Resolución del Consejo de 25 de junio de 2002 sobre «Conservar la memoria del mañana-Conservar los contenidos digitales para las generaciones futuras»(2002/C 162/02). DOCE C n. 162 (6.7.2002), p.4-5 (Disponible en: http://eur-lex.europa.eu/LexUriServ/site/es/oj/2002/c_162/c_16220020706es00040005.pdf). [Consulta: 5 de octubre de 2007].

8. Comisión de las Comunidades Europeas. *i2010. Bibliotecas digitales. Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: COM(2005) 465 final.* Bruselas, 30.9.2005 . (Disponible en: http://europa.eu.int/information_society/activities/digital_libraries/doc/communication/es_comm_digital_libraries.pdf). [Consulta: 5 de octubre de 2007].

cas europeas, tales como la Universidad de Oxford —desde el principio— y, recientemente la Biblioteca de Catalunya, con cuatro bibliotecas catalanas más, y la Universidad Complutense de Madrid⁹.

La **i2010: Bibliotecas Digitales**¹⁰ se ha configurado como uno de los proyectos nucleares dentro de la llamada *Estrategia i2010*, estrategia global de la Unión Europea en relación a la *economía digital*. Pues bien, en ella se definen las bibliotecas digitales como “colecciones organizadas de contenidos digitales que se ponen a disposición del público” y se señala ya que se trata tanto de objetos digitalizados como de aquéllos producidos en formato digital. Por tanto, la iniciativa “cubre ambos apartados: el material digitalizado y el digital desde su misma creación”.

Y en este sucinto repaso mencionaremos otra iniciativa internacional, ésta de corte profesional, en la que la IPA (International Publishers Association) y la IFLA (International Federation of Library Associations), firmaron en 2002 un documento de posicionamiento sobre la preservación del Patrimonio Digital: *Preserving the Memory of the World in perpetuity: a joint statement on the archiving and preserving of digital information*¹¹ en el que, entre otras recomendaciones se manifiesta que, dado que las bibliotecas nacionales tienen la misión de adquirir y conservar el patrimonio publicado en sus países respectivos, y dado que muchas de ellas ya están en la operación de incorporación de publicaciones digitales, esas bibliotecas junto con otros organismos y bibliotecas relevantes, deberían tomar el liderazgo en la responsabilidad de la conservación a largo plazo de las publicaciones digitales.

52

Las bibliotecas nacionales y el patrimonio digital

El punto anterior explica en buena medida el porqué son las bibliotecas nacionales las que se ocupan o, al menos, toman la iniciativa y llaman la atención sobre la importancia de los objetos digitales y sobre la necesidad de conservarlos para proteger su accesibilidad para las generaciones venideras, habida cuenta de que se trata de la memoria de nuestro tiempo.

Las bibliotecas nacionales se crean precisamente para conservar y proteger el patrimonio bibliográfico, en su momento, impreso y por eso los estados se dotan de leyes reguladoras de depósito legal obligatorio de la producción bibliográfica de un país. La legislación española, al igual que la de numerosos países europeos, obliga al depósito no solo de obras bibliográficas impresas sino también de mapas y planos, de grabaciones sonoras, de producciones cinematográficas, de obras en nuevos soportes como CDs, CD-IS, DVDs, etc.¹². Algunos países están

9. Disponible en: <http://www.google.es/press/pressrel/googlebooks.html>. [Consulta: 5 de octubre de 2007].

10. Disponible en: http://ec.europa.eu/information_society/activities/digital_libraries/what_is_dli/index_en.htm. [Consulta: 5 de octubre de 2007].

11. IFLA/IPA Steering Group. *Preserving the Memory of the World in Perpetuity: a joint statement on the archiving and preserving of digital information. Adopted by the IFLA/IPA Steering Group, meeting in Paris, 27th June 2002.* (Disponible en: <http://www.ifla.org/V/press/ifla-ipa02.htm>). [Consulta: 5 de octubre de 2007].

12. Disponible en: <http://www.bne.es/esp/bne/depositolegal.htm>. [Consulta: 5 de octubre de 2007].

desarrollando nuevas regulaciones del depósito legal para facilitar la conservación de documentos originalmente digitales, aunque la tendencia parece ser en la dirección de acuerdos voluntarios.

En el caso de los objetos digitales todas las iniciativas que se han puesto en marcha en diferentes países son iniciativas que se llevan a cabo entre las bibliotecas nacionales con la colaboración y/o la participación de distintos agentes. El caso más extremo sería el norteamericano en el que la Library of Congress lidera el macroproyecto *NDIIPP (National Digital Information Infrastructure and Preservation Program)* con 52 organismos en una red distribuida, del cual ella es la coordinadora e integradora, centrada en marcar la agenda estratégica facilitando los aspectos legales, técnicos, económicos, y llevando la iniciativa de la red pero sin aplicar un esquema de “arriba-abajo”¹³, hasta aquellos proyectos de dimensiones más limitadas como puede ser el caso catalán en el que también participan las universidades, como se verá más adelante.

Además de la definición de biblioteca nacional como institución de la memoria, converge también el hecho de que las bibliotecas incorporan las TIC desde muy temprana época, tanto para la facilitación de sus procesos técnicos como para la generación de servicios para los usuarios y puesta en valor de sus colecciones y de ahí el hecho de manifestar tan tempranamente esa inquietud por el futuro de los objetos digitales y ser proactivas respecto a su conservación. No obstante, lo que sí hay que remarcar es que sea quien sea la institución o entidad pionera, el proyecto de creación, mantenimiento y preservación del patrimonio digital será exitoso en la medida en que aquella sea capaz de entablar acuerdos y sinergias cooperadoras entre múltiples agentes, tanto pertenecientes a las administraciones públicas como a agentes de la esfera privada.

53

La colección digital

Como veremos en los casos que revisaremos un poco más adelante, es imprescindible definir la cobertura del proyecto a desarrollar, por qué modelo se va a optar en la recolección y captura de objetos digitales para su conservación y qué tipos de objetos se van a recoger y conservar.

En muchos casos, la incorporación de los tipos de objetos digitales no es una decisión cerrada sino que, en muchas ocasiones viene marcada por la tecnología y los recursos disponibles. Ahora bien, en la base de todos los proyectos que actualmente se desarrollan está la recogida, almacenamiento y preservación de los sitios web, aunque no sólo sea este el tipo de objeto digital. Por tanto, lo que sí es necesario definir es el “territorio virtual” de la colección que se pretende incluir en el proyecto.

En los casos cuyo liderazgo lo detentan las bibliotecas nacionales, el referente viene siendo la creación de las colecciones patrimoniales bibliográficas, es decir, lo que conocemos como *bibliografías nacionales*. El punto de partida es el dominio web que cubre una porción de ese

13. Disponible en: <http://www.digitalpreservation.gov/>. [Consulta: 5 de octubre de 2007].

territorio virtual fácilmente reconocible (.es para España, .uk para Reino Unido, .dk para Dinamarca, etc.), asignado por ICANN (*Internet Corporation for Assigned Names and Numbers*)¹⁴ y, por tanto, capturable, al igual que puede ser el criterio lingüístico, como el .cat para la comunidad lingüístico-cultural catalana (es interesante la estrategia de obtención de dicho dominio que se desarrolló a partir de la creación en el año 2001 de la Associació puntCAT¹⁵).

En la bibliografía revisada nos encontraremos con tres tipos de modelos a la hora de definir un proyecto de estas características: el *exhaustivo*, el *selectivo* y el modelo *híbrido*, es decir, intentar aprovechar lo mejor de los dos tipos. Nos valdremos de otras tantas experiencias para describir las características de cada uno de ellos.

a. Modelo exhaustivo / integral

Por *exhaustividad* se entiende la aplicación de robots o crawlers para la captura masiva automática, bien por dominios, bien por lengua, bien por área temática, etc. de la colección de documentos de la web.

Con ello se consigue una colección rica, contextualizada y con unos resultados visibles a corto plazo, además de suponer un coste más bajo que los paradigmas selectivos. Ahora bien, obtendremos una colección irregular en su composición que, dadas sus dimensiones, procedencias y propietarios, requerirá un acceso limitado.

54

Es el punto de partida de **Kulturarw3**¹⁶, el proyecto de preservación del patrimonio digital sueco, a iniciativa de la Kungl.Biblioteket¹⁷ (Biblioteca Real), la Biblioteca Nacional de Suecia, punto de partida que viene siendo común a la de otros países nórdicos como Noruega o Finlandia con los que tienen un foro de trabajo para coordinar sus actividades.

En el caso sueco que nos ocupa, iniciado en septiembre de 1996, la colección es **exhaustiva** y cubre revistas, periódicos, diarios electrónicos, ficheros estáticos diversos así como documentos dinámicos con enlaces y dado que son proyectos en constante evolución no está cerrado a la inclusión de otro tipo de objetos en el futuro. En la propia descripción del proyecto se afirma que en determinadas áreas, como la adquisición de objetos, tienen conocimiento claro de qué es lo que hay que hacer, al igual que tienen acceso a un equipamiento adecuado. Pero en otras áreas, como, por ejemplo en preservación a largo plazo, las decisiones se toman en función de las soluciones alternativas existentes en cada momento.

La captura de los documentos en la web se efectúa de manera automática mediante diversos robots de búsqueda. Estos robots realizan búsquedas sistemáticamente e indizan las páginas

14. Disponible en: <http://www.icann.org/tr/spanish.html>. [Consulta: 5 de octubre de 2007].

15. Disponible en: <http://www.puntcat.org/>. [Consulta: 5 de octubre de 2007].

16. Disponible en <http://www.kb.se/soka/internet/sv-webbsidor/>. [Consulta: 5 de octubre de 2007].

17. Disponible en: <http://www.kb.se/>. [Consulta: 5 de octubre de 2007].

encontradas por dominios “nacionales”, en este caso, .se. Un tercio de sitios web suecos están dados de alta en dominios internacionales como .com, .org, .net y otros y, mediante creación de algoritmos, desarrollan programas de búsqueda utilizados por estos robots.

El último barrido realizado en Kulturarw3 ha recogido unas 46 millones de ficheros URLs procedentes de 232.600 sitios web, por un volumen de 1,650 Tbytes. Los objetos, además de texto, imágenes, sonido, etc., son almacenados en cintas magnéticas en un robot de cintas (Tape Robot) y mediante un software de gestión de almacenamiento jerárquico es consultable semi-online, con un minuto aproximado de demora.

El total del patrimonio digital sueco almacenado asciende a unos 306 millones de URLs con un volumen de 10 Tbytes, de los cuales la colección de diarios asciende a 0,5 Tbytes y unos 6 millones de URLs.

Una cuestión de interés es la accesibilidad de toda esta colección que, dado que es actual, está sujeta a la propiedad intelectual y derechos de autor. La KB sólo puede permitir la consulta en ordenadores en sus propias instalaciones, de acuerdo con un decreto de 2002¹⁸ que reguló la accesibilidad a este patrimonio. Hasta entonces, y desde 1996, la KB había estado recogiendo todos los objetos digitales de la red dando por sentado que no estaba llevando a cabo otra labor que la estrictamente de conservación de un patrimonio tan efímero que si no se capturaba no iba a quedar ningún rastro en muy poco tiempo. En un momento dado, un particular puso en tela de juicio la legitimidad del proyecto y el Gobierno sueco publicó un decreto, encargando la misión de conservación del patrimonio digital a favor de la KB, pero contemplando la consulta estrictamente en sus dependencias físicas.

55

b. Modelo selectivo

En la aproximación selectiva, se limita la cobertura temática a la hora de crear la colección digital. La ventaja de este modelo sería el desarrollo de una colección más equilibrada, con un acceso más completo a la colección pero también una colección más descontextualizada que, necesariamente, proporcionará una visión limitada de la realidad actual.

Es un modelo con un coste más elevado que el anterior ya que aunque se ataca con medios tecnológicos en origen, posteriormente en la selección interviene el experto humano.

Es el diseño sobre el que se asentó PANDORA¹⁹, el Archivo Web de Australia, una creciente colección de publicaciones online australianas, más de 42 millones de ficheros en cifras de agosto de 2007, que comenzó la Biblioteca Nacional de Australia²⁰ en 1996, en cuanto se comenzaron a publicar documentos sólo digitales en la web, y ahora es un proyecto colabo-

18. Disponible en: http://www.kb.se/Info/Pressmed/Arkiv/2002/020605_eng.htm. [Consulta: 17 de mayo de 2005].

19. Disponible en: <http://pandora.nla.gov.au/apps/PandasDelivery/WebObjects/PandasDelivery.woa>. [Consulta: 5 de octubre de 2007].

20. Disponible en: <http://www.nla.gov.au/>. [Consulta: 5 de octubre de 2007].

rativo con otros nueve organismos culturales, entre ellos el National Film and Sound Archive²¹, de Australia.

Tal como se dice en su sitio web, PANDORA almacena **selectivamente** y proporciona acceso a **largo plazo** a sitios web y publicaciones online sobre Australia o de autor australiano en temas sociales, políticos, culturales, religiosos, científicos y económicos significativos y relevantes para Australia, o de autor australiano de reconocida autoridad que hace una contribución al conocimiento internacional.

Cada uno de los nueve organismos comprometidos en la recogida de objetos digitales tiene asignadas parcelas de selección de objetos digitales, según su misión y especialización.

En paralelo a las tareas técnicas de almacenamiento de colecciones digitales, PANDORA ha llevado a cabo tareas de investigación tecnológica, ha desarrollado modelos de gestión de la preservación digital, procedimientos de descripción de metadatos de preservación digital, etc. Aplicaciones desarrolladas para este proyecto se están utilizando para proyectos similares en otros países (p.e. software PANDAS para capturar sitios web).

c. Modelo híbrido

Vendría a ser una solución optimizadora de las dos visiones anteriores: por una parte, se recoge automática y exhaustivamente todo aquello que esté en una determinada lengua, en un determinado dominio, en servidores de una determinada zona geográfica, etc. y además habría una serie de tareas selectivas para incorporar objetos digitales temáticamente, algunos manualmente, otros estableciendo acuerdos de depósito voluntario con los agentes productores, sin perder de vista aquellos acontecimientos que tengan una relevancia especial en una determinada sociedad.

56

Para describir esta aproximación nos vamos a detener en el proyecto catalán, PADICAT²², *Patrimonio Digital de Catalunya*, que aunque es reciente en el contexto de iniciativas pioneras mundiales que se mencionaban, en España es la única iniciativa por el momento, que engarza con el resto de proyectos que se están llevando a cabo mundialmente en relación a la recogida y conservación del patrimonio digital. Además es un proyecto de referencia en Europa en lo que se da en llamar *web territorial*, es decir, depósitos digitales “transfronterizos”, temáticos, lingüísticos, etc.

Está encabezado por la Biblioteca de Catalunya, con el objetivo de “compilar, conservar y difundir la producción digital catalana”, entendiéndose por tal la creada en Catalunya, en lengua catalana, sobre Catalunya y/o de autoría catalana. Es decir, se hace extensivo al escenario digital la misión de conservación del patrimonio bibliográfico catalán en tanto depositaria del depósito legal.

21. Disponible en: <http://www.screenound.gov.au/>. [Consulta: 5 de octubre de 2007].

22. Disponible en: <http://www.padicat.cat/es/quees.php>. [Consulta: 5 de octubre de 2007].

PADICAT llevó a cabo durante el año 2005 el estudio de los proyectos ya existentes, así como los aspectos legales subyacentes, procediéndose ya a un test de hardware y software. Durante 2006 se puso en marcha la implementación del sistema y una fase piloto de captura y ajuste del proyecto.

PADICAT opta por una estrategia de recopilación de objetos digitales *exhaustiva* de todos los webs bajo el dominio .cat, webs en servidores de Catalunya, aquellos que se encuentren en lengua catalana en otros dominios, como .es, .com, .org, etc.

A día de hoy, si se visita el sitio PADICAT²³, se aprecia que la incorporación de webs de organismos e instituciones ha sido vertiginosa, contando ya con doscientos colaboradores. Dispone de 24 millones de ficheros, cifra que corresponde a 813 webs y 2.409 capturas.

La BC va a intentar formalizar acuerdos institucionales para conseguir el *depósito voluntario* y permanente de productores privados de contenidos digitales, como, por ejemplo, editoriales y otros (similar a como se está llevando a cabo en Países Bajos, en su proyecto DNEP²⁴, de depósito voluntario de publicaciones electrónicas).

Además de todo lo anterior, PADICAT también se compromete a una captura selectiva de acontecimientos relevantes para/en Catalunya y como muestra de ello se puede consultar el archivo de los web de las pasadas elecciones municipales en Catalunya el pasado mes de mayo de 2007²⁵.

PADICAT cuenta con el concurso de CESCA (Centre de Supercomputació de Catalunya) y el DURSI (Departament d'Universitats, Recerca y Societat de l'Informació. Generalitat de Catalunya). El presupuesto plurianual 2006-2008 contemplado asciende a 766.000€²⁶ (aprobado por Acuerdo de Gobierno de la GC, el 24 de enero de 2006), que incluye el 22% del total de la infraestructura tecnológica (el restante 78% lo pone el CESCA) y el equipo de recursos humanos encargado de PADICAT²⁷.

Las previsiones para el 2009 son de dotar al sistema con capacidad para 100.000 webs que comporten unos 50 millones de archivos por un volumen de datos de 30 Tbytes. Pretenden también para esa fecha tener firmados acuerdos con 300 entidades de todo tipo y poder dar acceso en línea a buena parte de la colección²⁸.

Aunque se podrían revisar muchos más casos de producción de depósitos nacionales digitales con otras problemáticas y otros agentes, básicamente, tendrían similitudes con los que aquí

23. Disponible en: <http://www.padicat.cat/>. [Consulta: 5 de octubre de 2007].

24. Disponible en: <http://webdoc.gwdg.de/edoc/aw/liber/lq-2-01/09Sijtsma.pdf>. [Consulta: 5 de octubre de 2007].

25. Disponible en: <http://www.padicat.cat/municipals2007.php>. [Consulta: 5 de octubre de 2007].

26. Disponible en: <http://www.genocat.net/acordsdegovern/20060124/02.htm>. [Consulta: 5 de octubre de 2007].

27. Disponible en: http://www.bnc.es/bc/in_padicat_2005.pdf. [Consulta: 5 de octubre de 2007].

28. Disponible en: <http://www.genocat.net/acordsdegovern/20060124/02.htm>. [Consulta: 5 de octubre de 2007].

hemos reseñado. Quien lo necesite por motivos profesionales o, simplemente, por curiosidad intelectual, encontrará una abundante bibliografía así como numerosos sitios web sobre el tema.

Accesibilidad del patrimonio digital

Término polívoco éste cuando de cuestiones documentales se trata. Vamos a distinguir aquí entre “acceso a la colección digital” en el presente, “accesibilidad” desde el punto de vista de la propiedad de los objetos digitales y “accesibilidad” futura a los objetos actuales, tema que se tratará al hablar de preservación.

El acceso a la colección, desde el punto de vista analítico-documental de técnicas de búsqueda y recuperación de información no se trata aquí ya que es un tema muy específico. Ahora bien, no es baladí el problema de la *accesibilidad* a la colección de objetos digitales habida cuenta de que son objetos producidos actualmente, amparados por la legislación de propiedad intelectual nacional y europea, objetos no regulados por ley de depósito legal en la mayoría de los casos, como mencionábamos al principio al hablar del patrimonio bibliográfico clásico.

Ya se mencionó el caso sueco en el que por decreto gubernamental se encargaba a la Kungl. Biblioteket de la recogida y conservación de la colección digital, así como el acceso exclusivo online desde sus instalaciones a esa colección digital, pero nunca acceso online ubicuo y múltiple.

58

Es por ello también que en numerosos proyectos se establecen acuerdos con terceros para incorporar al patrimonio digital colecciones de acceso cerrado como pueden ser las publicaciones científicas comerciales a las que únicamente se accede bajo suscripción, como en el caso holandés que, además, el país es sede de numerosas y fortísimas editoriales electrónicas como Elsevier, Kluwer, etc. y que a junio de 2000 habían depositado alrededor de 400.000 artículos científicos²⁹. Además de esto, y también sobre la base de acuerdos voluntarios, en el depósito electrónico de la Koninklijke Bibliotheek, la Biblioteca Nacional holandesa, tienen en marcha proyectos de depósito de publicaciones académicas de 16 instituciones universitarias de tesis, informes, artículos sobre la base de una red de “repositorios institucionales” que es como les denominamos en el argot profesional así como otro proyecto con 5 instituciones culturales para conservar copias maestras en formato .TIFF.

Hay pocas administraciones que hayan establecido regulatoriamente el depósito legal de objetos producidos en internet y quizás el más claro es el caso danés: el 22 de diciembre del 2004 se aprueba una nueva Ley de Depósito Legal³⁰ que, entre otros materiales, se refiere a

29. SIJTSMA, Lex. The Dutch Deposit of Electronic Publications. (DNEP) – 1995-2000. *Liber Quarterly*, 11, 2001, pp. 186-210. (Disponible en: <http://webdoc.gwdg.de/edoc/aw/liber/lq-2-01/09Sijtsma.pdf>. [Consulta: 5 de octubre de 2007].

30. Disponible en: <http://www.bs.dk/content.aspx?itemguid={4F101043-9A94-44A8-A6AA-CFD7819B6EC2}>. [Consulta: 5 de octubre de 2007].

materiales publicados en redes de comunicación electrónica y obliga a la entidad depositaria del depósito legal a su conservación.

El depósito legal danés establece que se considerará “danés” a efectos de conservación de patrimonio digital aquellos objetos *publicados en dominios de internet asignados a Dinamarca o publicados en otros dominios de internet dirigidos a públicos en Dinamarca*. Establece en quién recae la obligación de depositar la copia legal que de no ser así será demandado por los tribunales daneses.

Exista una regulación de obligado cumplimiento o un acuerdo voluntario entre partes, lo cierto es que la accesibilidad respecto a objetos digitales contemporáneos conllevará necesariamente una serie de limitaciones que deberán tratarse y definirse en una iniciativa de estas características.

Preservación del patrimonio digital

La *accesibilidad futura* a la producción electrónica es una de las variables más débiles en los proyectos de patrimonio digital y, paradójicamente es por lo que realmente surge esta inquietud por su conservación. Hay que recordar que el UKWAC (Web Archiving Consortium británico) citaba una vida media por sitio web de 44 días³¹, así que, es un documento efímero en unos formatos permanentemente cambiantes y soportados en unas aplicaciones en constante revisión.

Quiere esto decir que una vez iniciada y desarrollada la colección digital no está resuelto el problema. Si tenemos un manuscrito en pergamino necesitaremos unas condiciones de conservación del mismo, pero mucho más “controlables” que las necesarias para la conservación del patrimonio digital. En el primer caso, las garantías de permanencia son de cientos de años con un cierto control de la temperatura y humedad constantes; en el segundo, no hay garantías de permanencia y los esfuerzos dedicados son enormes, a sabiendas de que no hay soluciones definitivas.

En preservación digital se postulan tres aproximaciones que, sin ánimo de entrar en detalle, serían:

- a) *Preservación de la tecnología* que, básicamente, consiste en la conservación del objeto digital junto con el hardware y software necesario para su lectura y/o interpretación, es decir, sería tener finalmente un “museo” ilimitado de cachivaches para poder descifrar todos y cada uno de los objetos y es, por tanto, un esquema poco operativo.
- b) *Emulación de la tecnología*, es decir, preservar las cadenas de caracteres y el software utilizando unos programas emuladores, o simuladores, que se comportarían como el hardware obsoleto.

31. Disponible en: <http://info.webarchive.org.uk/pressrelease21-06-04.html>. [Consulta: 5 de octubre de 2007].

- c) *Migración de la información*, consistente en ir transfiriendo la información digital de una configuración a otra, centrada, pues en el contenido de los objetos digitales.

La migración es la estrategia de preservación más utilizada en proyectos de patrimonio digital, siendo así en el plan de PADICAT, Kulturarw3, etc.

Otra cuestión a tener en cuenta en el área de preservación será el almacenamiento de la colección. En algunos proyectos está prevista la “ubicación” de dos copias físicas, alojadas en ordenadores de dos edificios distintos, de dos instituciones diferentes como sería el caso catalán.

En los países escandinavos, se generan cintas de ordenador, una especie de archivo de cintas, de jukebox de almacenamiento en masa, como copias de seguridad. Este archivo está dotado de una aplicación que controla los datos almacenados y se encarga de hacer copias de aquellas cintas que estén en condiciones no del todo seguras.

Para terminar...

A lo largo de estas páginas se ha tratado de poner de manifiesto, de modo divulgativo, la importancia, la complejidad y la pertinencia de la puesta en marcha de proyectos de recogida y conservación de documentos/objetos de naturaleza exclusivamente digital.

60

Aspectos sumamente nucleares en estos proyectos, no mencionados aquí por su especificidad, son no sólo los recursos físicos y las características del software a utilizar, sino los estándares y protocolos a aplicar tanto en la captura e identificación de los objetos digitales, cuanto en la descripción de los mismos, así como los estándares de compresión de la información, etc.

Tampoco son menores ni menos relevantes los debates colindantes que están teniendo lugar en el ámbito académico y científico con los editores de publicaciones científicas electrónicas, el empuje del *acceso abierto* surgido en torno a la llamada *Iniciativa de Budapest*³² en diciembre de 2001 y los esfuerzos, también en España, por la creación de *repositorios institucionales* dentro de las iniciativas de acceso abierto (OAI) que, por un lado, ofrezcan visibilidad a los trabajos científicos y, paralelamente, se puedan conservar para el futuro desarrollo y estudio de la ciencia.

Todas estas iniciativas son convergentes con las políticas europeas de utilización de los contenidos y recursos electrónicos para el desarrollo cultural y social y la preservación de la memoria y, por tanto, el desarrollo de lazos estratégicos en esa dirección contribuye a la condición de posibilidad de la existencia de una sociedad en el espacio electrónico, hoy por hoy todavía limitado, pero en poco tiempo indudablemente dominante.

32. Disponible en: <http://www.soros.org/openaccess/esp/read.shtml>. [Consulta: 5 de octubre de 2007].