

ARCHIVO WEB DE LA PUBLICACIONES EN LÍNEA EN LAS COMUNIDADES AUTÓNOMAS

MARÍA JESÚS MOLINA SUÁREZ*

Fecha recepción: 23 de marzo de 2018
Fecha de aceptación: 13 de abril de 2018

*Sin bibliotecas, ¿qué nos quedaría?
No tendríamos pasado ni futuro*
Ray Bradbury

I INTRODUCCIÓN Y ANTECEDENTES

La Biblioteca Nacional de España denomina «archivo web» en su página <http://www.bne.es/es/Colecciones/ArchivoWeb/index.html>, a la colección formada por la recolección automatizada de sitios web. Un archivo web está compuesto por páginas y sitios web cuyos contenidos han sido concebidos para su publicación en redes de comunicación. La finalidad de un archivo web es la preservación y difusión de estos recursos «nacidos digitales» para que puedan servir como herramienta de conocimiento para generaciones presentes y futuras.

Desde que en 1996 Internet Archive se propuso crear la gran biblioteca de Internet y empezó a guardar webs de todo el mundo, muchas instituciones —principalmente archivos y bibliotecas

* Biblioteca Pública del Estado de Las Palmas. Calle Muelle Las Palmas, s/n. 35003 Las Palmas de Gran Canaria.

nacionales— la han seguido. La Biblioteca del Congreso de los Estados Unidos, la Biblioteca Británica, las bibliotecas nacionales de Australia, Canadá, Francia, Suiza, Suecia, Dinamarca, Japón y muchas otras han estado rastreando y guardando la web.

En 2003 la Unesco publicó las *Directrices para la preservación del patrimonio digital*, en las que instaba a las instituciones patrimoniales a actuar inmediatamente para impedir que el patrimonio documental existente en la red de redes se perdiera. Muchas de las bibliotecas nacionales se integraron en el Consorcio Internacional para la Preservación de Internet (IIPC), fundado en 2003 y que aglutina organizaciones de más de veinte y cinco países, incluyendo bibliotecas y archivos nacionales, regionales y universitarios. Ejemplo de ello son la Library of Congress, la Bibliothèque Nationale de Francia, la British Library o la Biblioteca Nacional de España (BNE), que se integró en el Consorcio en el año 2010.

En 2006 se publica en *Diario oficial* L236 de 31/8/2006, la recomendación de la Comisión Europea de 2 de agosto de 2006 sobre la digitalización y la accesibilidad en línea del material cultural y la conservación digital. El soporte digital permite el acceso al material cultural europeo, que incluye documentos impresos (libros, revistas y periódicos), fotografías, patrimonio cinematográfico, objetos de museo, documentos de archivo o material audiovisual. Las principales ventajas de la digitalización en línea del material cultural pueden resumirse así:

- ampliar su acceso a gran parte de la población de la Unión Europea (UE);
- dar un perfil claro al patrimonio diverso y multilingüe de la UE;
- conservar a largo plazo, en beneficio de las generaciones futuras, esta memoria colectiva.

La comisión pide la coordinación necesaria entre los estados miembros y hace hincapié en la «recolección en la web», una nueva técnica de recogida de material en línea con fines de conser-

vación. Consiste en que determinadas instituciones con un mandato específico recojan material en lugar de esperar a que éste sea depositado. La comisión propone integrar ese procedimiento en las legislaciones nacionales, con lo que se reducirá la carga administrativa de los productores de material digital. Recomienda además, una colaboración eficaz entre los estados miembros para evitar la variedad actual de disposiciones nacionales que regulan el depósito de material en un entorno digital.

En 2011, la Comisión Europea adopta un recomendación que publica en un comunicado de prensa: «Agenda Digital: digitalización del patrimonio cultural de la UE para impulsar el crecimiento», en el que tiene como punto de partida las conclusiones del informe del Comité de Sabios (un grupo de reflexión de alto nivel) sobre la digitalización del patrimonio cultural europeo, elaborado en 2010. La adopción de medidas en apoyo de las industrias culturales y creativas y la consecución de un modelo sostenible de financiación de Europeaana figuran entre los objetivos de la Agenda Digital para Europa.

2 DEPÓSITO LEGAL DE PUBLICACIONES EN LÍNEA

Actualmente, todo lo relacionado con el llamado «Archivo de la web», se traslada a un nuevo concepto «Depósito Legal electrónico o de publicaciones en línea» que engloba lo anterior. Por citar algunos ejemplos de países donde se ha legislado ya al respecto, destacan la Biblioteca Nacional de Francia (Code du Patrimoine, 2006) y la Biblioteca Nacional de Reino Unido (Non-print Legal Deposit, 2013).

En España, para cumplir con el real decreto aprobado el 10 de julio de 2015 por el Consejo de Ministros que regula el depósito legal de las publicaciones en línea y que desarrolla la *Ley 23/2011, de 29 de julio, de Depósito legal*, en la que se consideran por primera vez objeto de depósito legal los sitios web y las publicaciones en línea, tendremos que implementar un proceso tec-

nológico para capturar, organizar, desarrollar, preservar y difundir todo el material publicado en la red de redes.

Según esta ley, son objeto de depósito legal todo tipo de sitios web y las publicaciones en ellos contenidas, siempre que contengan patrimonio bibliográfico, sonoro, visual, audiovisual o digital de las culturas de España. Además de regular los procedimientos de constitución del depósito de los sitios web y de los recursos que incluye, tiene como uno de sus principales objetivos el «permitir el acceso a los usuarios», especialmente con «fines de investigación», a los recursos distribuidos a través de las redes electrónicas.

Además, serán objeto de depósito legal las que estén producidas o editadas por cualquier persona o entidad que tenga su domicilio, residencia o establecimiento permanente en España; o que estén producidas o editadas bajo un nombre de dominio vinculado al territorio español.

Dadas las especiales características de este tipo de publicaciones y la imposibilidad de ser exhaustivos en su captura, almacenamiento y preservación en las condiciones tecnológicas actuales, el texto legal introduce una serie de novedades con respecto a la regulación del depósito legal de las publicaciones en soporte tangible:

—La iniciativa del depósito no está en los editores, sino en los centros de conservación (la Biblioteca Nacional de España y los designados por las CC.AA. en el ámbito de sus competencias), que determinarán las prioridades en el patrimonio documental en línea a preservar.

—No se asignará número de depósito legal a las publicaciones en línea. Dadas las especiales características de las publicaciones en línea y los límites difusos de los documentos difundidos por esta vía, la asignación de número de depósito legal excedería la capacidad de las oficinas de depósito legal de las CC.AA., que son las encargadas de asignarlo para las publicaciones en soporte tangible. Se simplifica así el procedimiento de gestión. Por otra parte, no parece necesario, ya que existen números normalizados, como el ISBN, el ISSN o el DOI, para la identificación unívoca de las publicaciones.

Para el ingreso de estas publicaciones en los centros de conservación como depósito legal, se prevén varias vías:

—La captura automática por medio de robots que recolectan la web, para aquellos sitios web y publicaciones contenidas en ellos que estén libremente accesibles en Internet.

—En el caso de publicaciones protegidas con usuario y contraseña o no capturables automáticamente (bases de datos, *e-books...*), los editores y los centros de conservación acordarán la forma más eficaz y sencilla para ambas partes de hacer el depósito.

El acceso a las publicaciones ingresadas por esta vía respetará escrupulosamente la legislación en materia de propiedad intelectual y de protección de datos de carácter personal.

La obligación de los editores y productores de contenido en línea se limita a dejar que los centros de conservación recolecten sus publicaciones o a facilitar el depósito cuando éstos se lo soliciten. Los centros de conservación se pondrán en contacto con los editores para decidir juntos la forma más eficaz y sencilla de transferir los contenidos, como la descarga vía FTP o la transferencia de los archivos en un dispositivo de almacenamiento, cuando las publicaciones no estén accesibles libremente en la web.

Se trata, pues, de un medio que garantiza la conservación de toda la producción editorial y debe ser contemplado como un bien democrático común que asegura a los ciudadanos de hoy y a las generaciones futuras el acceso al legado intelectual con fines de información e investigación.

EL ARCHIVO WEB EN LA BIBLIOTECA NACIONAL DE ESPAÑA

La Biblioteca Nacional de España lleva rastreando y guardando el dominio *.es* desde el otoño de 2009, con la ayuda de Internet Archive. Desde entonces se han llevado a cabo ocho recolecciones masivas del dominio *.es* y cuarenta y cinco recolecciones sobre

distintos eventos y materias previamente seleccionados, como complemento de las primeras. En el momento de redactarse este informe se está llevando a cabo la recolección de las páginas web de las elecciones generales de este año, aún inconcluso. De manera adicional, se ha realizado también el depósito a través de FTP o Protocolo de Transferencia de Archivos, de publicaciones cuya naturaleza impedía su recolección automática. Todas ellas constituyen una colección web que asciende a 130'63 TB de información comprimida y en la que la biblioteca trabaja para poder dar acceso a ella lo antes posible.

La cantidad de documentación en línea que hay que preservar es tan ingente que es necesario apoyarse en los sistemas que rastrean la web de manera automática y periódica, y asumir que —en el estado actual de la tecnología— guardarlo todo será imposible de momento. De ahí que el objetivo de esta regulación sea permitir a los centros de conservación obtener una muestra lo más amplia y representativa posible de toda la documentación disponible en Internet.

Hay un aspecto muy nuevo en este real decreto: el sujeto depositante, el obligado al depósito legal, no tiene que depositar nada. Serán los centros conservadores (los designados como tales por las comunidades autónomas y la Biblioteca Nacional de España) los responsables de rastrear los sitios web y archivarlos. Los responsables de esos contenidos en línea sólo deberán dejarse rastrear. De esta manera, se trata de facilitar a todas las partes implicadas la constitución del depósito legal. Para aquellos sitios web a cuyos contenidos sólo se pueda acceder mediante usuario y contraseña, se establece que los titulares deberán facilitar a los centros conservadores las claves necesarias para poder acceder a los contenidos, rastrearlos y guardarlos, exclusivamente con el fin de preservar el patrimonio documental y dar acceso a él en las mismas condiciones en las que son accesibles hoy en los centros de conservación todos aquellos fondos —en soporte tangible— que han ingresado por depósito legal, es decir, tomando las medidas necesarias que hagan compatibles el acceso a la información

y el respeto a la legislación en materia de propiedad intelectual y de protección de datos personales.

Es la obligación de los centros patrimoniales preservar el patrimonio documental para ponerlo a disposición de las generaciones futuras. Y esto sólo será posible con la colaboración de todos los sectores implicados. Los centros de conservación cuentan con personal especializado bibliotecario y documentalista para cumplir con esta misión: surge así el perfil profesional de los conservadores web.

El cumplimiento de este real decreto no debería implicar gasto alguno. De acuerdo con la declaración conjunta de la Conferencia de Bibliotecas Nacionales Europeas y de la Federación de Editores Europeos, se incluye en el real decreto una formulación que garantice que el cumplimiento de la ley no *«perjudique los legítimos intereses de los que detentan los derechos ni entre en conflicto con la explotación comercial que ellos hagan de dicho material»* (art. 8.4). Asimismo, se especifica que *«En ningún caso el cumplimiento de lo dispuesto en este artículo podrá suponer a los editores o productores una carga económica adicional a la directamente derivada de la mera transferencia de los contenidos y sitios web que permita dar cumplimiento a la obligación legal de depósito de las publicaciones en línea»*.

3 METODOLOGÍA

La tecnología utilizada por la Biblioteca Nacional de España en colaboración con los centros conservadores designados en cada comunidad autónoma para capturar «semillas» es la siguiente:

Las recolecciones de la web (*web crawls* en inglés) se llevan a cabo con robots de rastreo (*web crawlers*), que recorren la web a partir de una serie de URL iniciales llamadas *seeds* o semillas. Los *crawlers* toman estas URL como punto de partida para comenzar la recolección, y van guardando de forma automática cada página o sitio identificado, recorriendo todos los hiperenlaces que encuen-

tran y añadiéndolos a la lista de URL que visitarán recursivamente, almacenando de forma ordenada los documentos, vídeos, imágenes, etc. que se ocultan tras cada vínculo. Estos contenidos son indexados posteriormente con el fin de poder realizar búsquedas a texto completo en la colección.

Heritrix es un programa de software diseñado por el Internet Archive para el rastreo y captura de páginas y sitios web. Heritrix fue diseñado para minimizar el impacto perturbador de las recolecciones sobre la actividad normal de los sitios web: el ritmo de rastreo del programa es adaptable y puede ajustarse según parámetros escogidos para tal fin. La versión actual de Heritrix utilizada es la v.3.1.1. Para llevar a cabo todas las acciones relativas a la recolección automática del dominio .es, el robot rastreador se identifica con el nombre `archive.org_bot`.

Esta identidad es distinta a la del robot que habitualmente se asocia con el Alexa Internet Archive, designado como `ia_archiver`. El robot recolector está programado para respetar las recomendaciones de los ficheros tipo `robots.txt` y de las etiquetas `robots META tags`, según lo especificado en el sitio web `robotstxt.org`. En este mismo *site* se proporciona también información sobre aquellos casos excepcionales en los que para las búsquedas se empleará el agente especial de rastreo `img+css-bot` y la referencia (*referrer*) mostrará la URL contenida en los registros. Para detener el robot rastreador, se debe especificar el agente mencionado en el párrafo anterior y modificar el fichero `robots.txt`, como en el ejemplo que figura a continuación:

```
User-agent: archive.org_bot  
Disallow: /
```

Hay que tener en cuenta que el robot rastreador volverá a comprobar el fichero `robots.txt` cada veinte y cuatro horas. Si el rastreo se quiere detener antes, hay que enviar un mensaje al buzón del Internet Archive o al equipo de la Biblioteca Nacional de España encargado del proyecto.

NAS (NetarchiveSuite) es el paquete de herramientas que se usa para manejar el robot. Desde él se programan las recolecciones y se determina la profundidad de cada semilla y la frecuencia con que debe lanzarse cada recolección. La usan para el archivo web la biblioteca nacional de Dinamarca, la de Francia y la de Austria.

Hasta finales de 2015 las recolecciones web se habían llevado a cabo en una versión de NAS, la 4.2, que en realidad era un entorno de pruebas por la infraestructura en la que estaba montado, pero ha servido para hacer recolecciones reales (aunque no de mucho alcance). A finales de 2015 se instaló un entorno de producción de NAS, la versión 4.4, sobre la que se están realizando las recolecciones ahora mismo. El entorno 4.2 se está «vaciando» para convertirlo en un entorno de pruebas, conectado a cweb de reproducción, en el que se puedan hacer pruebas sin interferir con las recolecciones en producción.

cweb ('colecciones en la web') es una herramienta que supone una capa por encima de NAS para que los conservadores web (tanto de la BNE como de las CC.AA.) puedan manejar desde ahí sus colecciones. Esta herramienta la ha desarrollado la Biblioteca Nacional de Francia y la ha cedido a la BNE en virtud de un convenio de colaboración entre las dos instituciones y del marco de colaboración del Consorcio Internacional para la Preservación de Internet (IIPC), del que ambas instituciones son miembros. Permite a los conservadores web seleccionar y gestionar las colecciones de sitios web, introduciendo datos como URL, parámetros de recolección (frecuencia, profundidad y tamaño) y descripción (materia, palabras clave, notas de contenido y notas técnicas). Las listas de los sitios seleccionados se transfieren al robot rastreador de NetArchiveSuite (NAS) para ser recolectados, con la configuración que se haya especificado en cweb. Los resultados de la recolección se pueden ver en OpenWayback. cweb no proporciona acceso a los sitios archivados.

OpenWayback es el software que permite recuperar lo recolectado, mediante URL. Hay que conocer la URL exacta. De mo-

BNE
BIBLIOTECA NACIONAL DE ESPAÑA

Depósito Legal Electrónico

Archivo Web

A través de esta página se ofrece a los conservadores web un acceso provisional a los resultados de las recolecciones web, tanto masivas como selectivas, llevadas a cabo desde 2009

- OpenWayback** Consulta de las colecciones archivadas
- Pre-CWeb** Acceso al entorno de Preproducción de CWeb

ADVERTENCIA: Los contenidos a los que se accede a través de esta interfaz están protegidos por la Ley de Propiedad Intelectual, aunque fueran de acceso libre cuando se archivaron. Por ello, no pueden copiarse, descargarse o difundirse por ningún medio, sin el consentimiento expreso de los titulares de los derechos.

Más información: archivoweb@bne.es

mento es el único sistema de recuperación del que disponen. Los informáticos trabajan en la indización a texto completo de todo lo recolectado para poder buscar por palabras, pero la cantidad de datos que hay que procesar es muy grande y no es fácil. Otras formas de recuperación en las que se está trabajando son por colecciones y por títulos. Esto requiere también cierto nivel de descripción bibliográfica (catalogación).

Www.archivoweb.es es la interfaz que han creado para que los conservadores web puedan acceder a las herramientas de colaboración disponibles. Poco a poco irán añadiendo nuevos entornos, algunos se modificarán y es probable que sea desde ahí desde donde den acceso a los usuarios para consultar el archivo web y las publicaciones en línea que ya están ingresando por depósito legal. Ahora mismo desde esa interfaz se da acceso a:

—OpenWayback. Recolección elecciones generales: esto es el entorno NAS 4.4. y muestra lo recolectado en él, que es aún muy poco ya que sólo funciona desde finales de 2015. El nombre es orientativo, porque en ese entorno ya se están lanzando otras colecciones. Así que el nombre cambiará.

—OpenWayback. Recolecciones selectivas anteriores: esto lleva a lo recolectado con el entorno NAS 4.2.

—Cweb: de momento, como decía antes, este entorno de cweb no se comunica con NAS, es decir, no sirve para recolectar. Sirve para probar la herramienta.

Existe otro contenedor de recolecciones (digamos otra Open-Wayback), que es el que contiene el resultado de las recolecciones masivas que Internet Archive hizo para la BNE desde 2009 a 2013.

Tienen hasta ahora, tres contenedores de información archivada:

1. El procedente de Internet Archive (no accesible)
2. El que contiene las recolecciones llevadas a cabo por la BNE desde 2014 en el entorno de NAS 4.2
3. El que contiene las recolecciones llevadas a cabo por la BNE desde finales de 2015 en el entorno de NAS 4.4

A día de hoy es imposible aspirar a la exhaustividad en el archivado web, debido al enorme tamaño de Internet y a los medios tecnológicos de los que disponemos actualmente. Por eso, para intentar guardar la mayor cantidad posible de información web, las bibliotecas han optado por un modelo mixto que combina recolecciones masivas y selectivas.

Las recolecciones masivas recogen el mayor número de dominios posible con una profundidad pequeña en los niveles de navegación (inicio o inicio+1). El objetivo es recoger una muestra representativa del panorama de la web. Para ello se sacrifica la profundidad en beneficio de la exhaustividad. Para cumplir con el real decreto del depósito legal de publicaciones en línea, la Biblioteca Nacional de España hará una recolección anual del dominio .es sobre la base del listado completo de todos los dominios registrados en ESNIC (1 600 000 sitios en 2013).

Las recolecciones selectivas se hacen para completar las recolecciones masivas y recogen con mayor profundidad y frecuencia una muestra más pequeña de sitios web seleccionados por su relevancia para la historia, la sociedad y la cultura. Para cumplir con el real decreto del depósito legal de publicaciones en línea, estas recolecciones se realizarán, en principio, en la Biblioteca Nacional de España con la colaboración de los centros de conservación de las comunidades autónomas, también depositarios del depósito

legal de publicaciones en línea, y de otras instituciones especializadas en distintas áreas del conocimiento, excepto en el caso de Cataluña y País Vasco que desde hace algunos años tienen sus propios sistemas de recolección web.

Las recolecciones selectivas pueden ser de tres tipos:

1. Temáticas: para capturar una selección de recursos sobre un tema o tipo de documento.
2. Evento: sobre acontecimientos de especial relevancia para la sociedad española.
3. Emergencia: en casos que impliquen riesgo de pérdida del sitio web, por ejemplo por cierre de instituciones, como la Agencia de Protección de Datos de la Comunidad de Madrid, Radio Televisión Valenciana, el Archivo Histórico Minero, etc.

Los criterios de selección que se exponen a continuación son un marco de referencia para los profesionales de las bibliotecas, están abiertos al debate y son susceptibles de revisión. Su aplicación tiene una parte subjetiva, inherente a todo proceso de selección, por lo que se recomienda que la labor de selección sea realizada por más de una persona.

Partimos de unos criterios generales de selección basados en el artículo 3 del real decreto que regula el depósito legal de publicaciones en línea, según el cual los sitios web objeto de depósito legal son:

- Sitios web que contengan patrimonio bibliográfico, sonoro, visual, audiovisual o digital de las culturas de España.
- Sitios web del dominio .es y subdominios asociados.
- Sitios web de otros dominios del territorio nacional.
- Contenidos españoles albergados en otros dominios (.com, .net, .edu, .org, etc.).
- Sitios web en cualquiera de las lenguas españolas oficiales.
- Sitios web de todo tipo, en cualquier formato, incluyendo las publicaciones en ellos contenidas.

—Sitios web tanto de acceso libre como restringido. En este último caso es necesario ponerse en contacto con el proveedor para que proporcione las claves de acceso y el robot pueda entrar a recolectar, o habilitar la manera para que puedan transferir esos contenidos por alguna vía, sea FTP u otro medio.

Los encargados de realizar la selección son los conservadores web. Éstos son bibliotecarios o documentalistas especializados que se encargan de identificar y seleccionar los recursos en línea objeto de depósito legal para recolectarlos y guardarlos en un archivo web.

Por último, existen varias categorías de recursos web que es recomendable incluir en las selecciones de sitios web a recolectar:

1. Medios de comunicación social: cabeceras de prensa, agencias de noticias, cadenas de radio y Televisión...
2. Organismos administrativos: ministerios, comunidades autónomas, ayuntamientos...
3. Instituciones políticas: partidos políticos...
4. Instituciones culturales: museos, archivos, bibliotecas, universidades, centros de investigación...
5. Asociaciones: partidos políticos, ONG...
6. Blogs y páginas web de personas relevantes relacionadas con el tema.
7. Redes sociales: Twitter, Facebook...
8. Wikis: Wikipedia...
9. Grabaciones en vídeo: YouTube...

CREACIÓN DE UN ARCHIVO WEB EN LA COMUNIDAD AUTÓNOMA DE CANARIAS

En la Comunidad Autónoma de Canarias se designa como centro conservador a la Biblioteca Pública del Estado en Las Palmas para que constituya repositorios para conservar y difundir las pu-

blicaciones en línea del patrimonio digital canario que vayan a capturar. El Real Decreto 635/2015 enumera los siguientes actos que podrán llevarse a cabo por los centros de conservación, con los requisitos que establece el art. 37.1 del Texto Refundido de la Ley de Propiedad Intelectual: *«los centros de conservación podrán reproducir, reformatear, regenerar y transferir los archivos para garantizar su conservación, respetando la legislación sobre propiedad intelectual»*.

En los supuestos en que los repositorios incluyan recursos de acceso restringido queda claro que, *«la consulta de los usuarios se llevará a cabo únicamente desde los terminales instalados en las sedes de la Biblioteca Nacional de España y de los centros de conservación de las comunidades autónomas, respetando la legislación sobre protección de datos de carácter personal y propiedad intelectual»*.

En este centro se nombrará a más de una persona especializada que serán denominadas conservadores web. Los conservadores web tendrán acceso al módulo de usuario de cweb, a través de la red SARA. El acceso a cweb está garantizado por el protocolo HTTPS para los usuarios externos a la BNE. cweb es compatible con los navegadores Firefox (versión 6 o superior) e Internet Explorer (versión 8 o superior). Algunas funciones de la aplicación solamente funcionan con los navegadores Firefox. Las páginas del sitio web han sido optimizadas para una resolución de pantalla de 1024 x 768. cweb requiere tener habilitadas las cookies y JavaScript.

La duración de una sesión en cweb sin actividad se limita a sesenta minutos. Después de esos sesenta minutos, la sesión se cierra automáticamente.

Para añadir o modificar sitios en cweb es necesario conectarse introduciendo nombre de usuario y contraseña en la pantalla de inicio de cweb. Para ello hay que estar dado de alta como usuario (si no, sólo están disponibles las opciones de consulta). Los conservadores web que no pertenezcan a la BNE enviarán una solicitud de acceso a archivoweb@bne.es indicando: Nombre y apellidos; Comunidad autónoma; DNI; Email. Cuando estén dados

The screenshot shows the 'Colecciones en la Web' interface. At the top, there is a BNE logo and a search bar with the text 'Buscar un sitio web' and an 'OK' button. Below the search bar, the text 'Entorno de Reproducción' and 'Búsqueda avanzada' are visible. The main content area is divided into three tabs: 'Colecciones estables (17)', 'Colecciones de eventos (2)', and 'Colecciones inactivas (4)'. The 'Colecciones estables' tab is active, showing a list of regions and the number of websites associated with each: Andalucía (36 sitio web), Aragón (1 sitio web), Asturias (25 sitio web), Baleares (0 sitio web), Canarias (6 sitio web), Cantabria (7 sitio web), Castilla La Mancha (15 sitio web), Castilla y León (4 sitio web), Extremadura (18 sitio web), Fotografía (20 sitio web), and Galicia (15 sitio web). On the right side, there is a sidebar with a login section titled '¿Ya tienes una cuenta?' containing fields for 'Nombre de usuario' and 'Contraseña', a 'Conectarse' button, and a message 'Nombre de usuario o contraseña incorrectos'. Below the login section, there is an 'Información' section with the text 'En este momento hay 466 sitios web seleccionados' and 'Archivo de la Web Española'. At the bottom of the sidebar, there is a '¿Preguntas? ¿Comentarios?' section with contact information: '91 516 81 18' and 'archivoweb@bne.es'. The footer of the page contains the text '© BnF 2016 - © BNE 2016 | CWeb version 5.3.0'.

de alta recibirán un *email* informando y pidiendo que cambien la contraseña. Después recibirán otro *email* avisando de que ya pueden acceder a cweb, si han cambiado la contraseña.

El código para el nombre de colecciones en caso de Canarias es CNR.

—CICLO DE VIDA DE UN ARCHIVO WEB

Este informe está basado en «El modelo de ciclo de vida del archivado web» descrito por Kristine Hanna, jefa del servicio de archivado de Internet Archive, en *Anuario A/CE de cultura general*, 2014, ya que su desarrollo se basa en la experiencia y en el aprendizaje adquirido por el flujo del trabajo de distintas organizaciones que conservan la red.

El fin es desarrollar buenas prácticas y metodologías a la hora de cumplir los objetivos, cuya dificultad se debe en parte a la evolución constante de la tecnología web, ya que puede resultar complicado archivar determinados tipos de contenidos de manera eficaz e incluso la adopción de políticas contradictorias y sujetas a cambios por parte de diversas partes interesadas, y los cambios en las estructuras organizativas y responsabilidades laborales también constituyen un obstáculo adicional a la hora de establecer buenas prácticas.

En el caso de Canarias como centro de conservación se debe a la innovación del trabajo y a la inexperiencia, por lo que con la aplicación de este modelo se tratará de constituir las fases por la que se ha de desarrollar la gestión del programa del archivado web.

El modelo constituye un intento de extraer los distintos pasos y fases por los que pasa una institución durante el desarrollo y la gestión de su programa de archivado web.

Aunque el modelo se desglosa en pasos individuales, cada acción no es independiente y éstos están relacionados.

1 Fases externas

Las fases siguientes forman parte del ciclo exterior del modelo, que está relacionado con las preguntas más amplias acerca de la creación y definición de un programa de archivado web institucional.

1.1 Misión, visión y objetivos

1.1.1 Misión

La misión del Centro Conservador Canario es capturar, conservar y difundir el patrimonio digital de Canarias nacido en Internet. El sistema se basa en la aplicación de una serie de programas infor-

máticos a los cuales tenemos acceso a través del Área de Gestión del Depósito de las Publicaciones en Línea de la Biblioteca Nacional de España, que permiten la captura, el almacenamiento, la organización y, aún en proceso de estudio, el acceso a las páginas web publicadas en Internet.

1.1.2 Visión

Hace años el Gobierno canario intentó proceder al registro del dominio .ic, como marca propia de Canarias en la red; sin embargo, la Secretaría de Estado de Telecomunicaciones no mostró ningún interés en dirigirse a la entidad Internet Corporation for Assigned Names and Numbers (ICANN) para atender esta petición. Según escrito en la prensa en 2012,

«ésta ha sido desde hace tiempo una reivindicación planteada desde Coalición Canaria (CC) con el fin de que las islas cuenten con un dominio propio dentro de Internet como reflejo de las supuestas especificidades y particularidades de las islas Sin embargo, ni el Estado ni el organismo que gestiona los dominios parece haber mostrado ningún interés ni urgencia a la hora de atender la demanda del ejecutivo autónomo. En respuesta a una pregunta presentada por el grupo parlamentario nacionalista canario en la Cámara regional, la Consejería de Presidencia se limita a señalar que "ante la falta de respuesta por parte de la Secretaría de Estado, se ha solicitado directamente a la ICANN que se asigne el dominio .ic". A partir de aquí "el organismo tendría que dirigirse al Ministerio de Industria español para recabar su no oposición al registro", indica la Consejería de Presidencia. Por ello, en estos momentos al gobierno canario no le queda más remedio que seguir a la espera de que tanto la ICANN como el Estado se pronuncien sobre la posibilidad de que las islas cuenten con un dominio propio en la red tal y como de forma insistente se viene demandando desde Canarias».

Cabe destacar en este apartado que la tienda Zara sí utiliza el código internacional ISO IC para su *store online* de Canarias, <http://www.zara.com/ic/>.

1.1.3 Objetivos

En esta etapa de nacimiento del proyecto, nos basaremos en recopilar la información digital que nos recomiendan en el artículo 3 del real decreto, antes descrito, aplicado a la cultura canaria, así como promover líneas de investigación procesando de manera monográfica los recursos de eventos de la vida pública canaria, como campañas electorales en Internet, prensa canaria, folklore o los museos canarios en Internet.

- a) Impulsar el depósito sistemático de la producción web de las entidades y las empresas de Canarias.
- b) Definir estrategias de preservación digital para el patrimonio canario nacido en Internet.
- c) Proporcionar radiografías periódicas de la web canaria.
- d) Detectar los formatos que experimentan a corto plazo problemas de ilegibilidad.
- e) Creación y mantenimiento de la hemeroteca digital en Internet, con la captura sistematizada de publicaciones digitales en serie.
- f) Cooperación con otros archivos web y depósitos de preservación digital, de bibliotecas, archivos y museos, para dar una respuesta eficiente a los retos de preservación digital y acceso a los recursos depositados.

1.2 Recursos y flujo de trabajo

En la Biblioteca Pública del Estado en Las Palmas hay ocho ayudantes de bibliotecas, de los cuales dos serán asignados como con-

servadores web para la creación del archivo web en este centro conservador. Al mismo tiempo, fundamentarán su trabajo en el *Manual para conservadores web*, versión 05, redactado por el Área de Gestión del Depósito de las Publicaciones en Línea de la BNE, y desarrollarán un «Manual de Procedimientos propio», aplicando las enseñanzas, las dudas, las creaciones y la experiencia que este trabajo les vaya determinando.

El presupuesto, en este período de creación y nacimiento del proyecto, es cero, ya que el personal responsable del archivo web es propio de la biblioteca y las herramientas y aplicaciones informáticas (antes mencionadas) las facilita la BNE.

Se determinará cuánto tiempo de su horario laboral dedicarán los dos conservadores web a examinar los rastreos o cuántos sitios web agregarán a la semana y la creación de un calendario para la realización de determinadas capturas selectivas.

1.2.1 Acceso, utilización y reutilización

Esta fase aún no está disponible para los usuarios, sólo para los conservadores web, que podrán comprobar si las sedes web que están capturando se recogen y se guardan de manera correcta.

En un futuro, se establecerán las políticas de acceso, utilización y reutilización para que el programa de archivado web tenga éxito. El Centro Conservador Canario planteará si permite el acceso abierto a sus archivos web y el modo de hacerlo, si promociona las colecciones y cómo hacerlo, además de cómo administrar el uso público del material.

Parte de la elaboración de una política de acceso comprenderá la elección de la tecnología o herramientas específicas para facilitar el acceso a las páginas web archivadas. En principio, utilizamos el software OpenWayback, una herramienta de visualización donde introduciremos la URL en la caja de búsqueda de OpenWayback; después seleccionamos una fecha de captura, y navegaremos por el sitio web archivado siguiendo un patrón similar al de la web viva.

1.2.2 Conservación

El director general de Red.es, Daniel Noguera, y la directora de la Biblioteca Nacional de España, Ana Santos Aramburo, han firmado el acuerdo por el que hasta 2018 se destinarán cinco millones de euros para conservar los fondos culturales y desarrollar el depósito legal de las publicaciones en línea, mediante el uso intensivo de las Tecnologías de la Información y las Comunicaciones.

«La cantidad de contenidos que se generan en soporte digital, lo efímero de su permanencia y su importancia como fuentes de información para el futuro obliga a asegurar su preservación y transmisión. El convenio firmado con Red.es garantiza los recursos suficientes para que la BNE continúe hoy con la misión que inspiró su fundación hace más de 300 años, la conservación de nuestra cultura, de aquello que nos identifica, independientemente del soporte en que se genere»,

ha destacado doña Ana Santos Aramburo.

1.2.3 Gestión del riesgo

El Centro Conservador Canario debe plantear cómo enfocar el riesgo a la hora de desarrollar un programa de archivado web, analizando los derechos de autor y los permisos, además del acceso. A la hora de desarrollar un programa de archivado web, debe considerar el nivel de riesgo relacionado con los derechos de autor que están dispuestos a aceptar y la manera en que gestionarán dicho riesgo, aunque estén protegidos por la Ley de Propiedad Intelectual.

2 Fases internas

Las fases restantes del modelo describen las actividades cotidianas propias de la gestión de un programa de archivado web.

Describe las tareas cotidianas que implica la actividad de archivar la web.

2.1 Valoración y selección

Aquí se decidirá qué sitios web serán recopilados en esta fase de inicio del proyecto. Básicamente producción digital canaria y relativa a las islas Canarias. Se comenzaría con una sección de Prensa Canaria, las URL de otros repositorios ya creados como Jable o la Memoria Digital de Canarias, instituciones públicas como Gobierno y cabildos, y Cultura Canaria, en la que incluiremos museos de arte, museos etnográficos, salas de arte, centros de investigación, etc.

2.2 Definición del alcance

Tras elegir qué sitios se van a archivar, se decidirá si se archivan completa o parcialmente. Esta labor puede realizarse antes de capturar la primera página o después de recopilar el contenido, como parte de una revisión general de la calidad de la recopilación. Esta parte del ciclo de vida puede ser bastante técnica en función de las herramientas que se utilizan.

2.3 Captura de datos

Aquí se ajustará el modo en que se capturan los datos decidiendo la frecuencia de rastreo y los tipos de ficheros que se archivarán o no. Las fases de definición del alcance y de captura de datos del ciclo de vida suelen solaparse porque conllevan actividades y decisiones similares. Una vez que se han decidido los sitios web que se van a capturar y el modo de hacerlo, se definirán la frecuencia y los tiempos de rastreo.

2.4 Almacenamiento y organización

Este paso comprende un plan de almacenamiento provisional o a largo plazo de los sitios web archivados. La fase de almacenamiento y organización del ciclo de vida también puede constituir sus actividades de conservación. Una vez que las instituciones han capturado los datos de los sitios web, deben examinar lo que han archivado y evaluar su calidad e integridad. Esta tarea puede realizarse mediante informes generados por los rastreadores o examinando los propios archivos mediante una herramienta de acceso como el software OpenWayback.

2.5 Aseguramiento y análisis de la calidad

Se examinará el material archivado y en qué medida la colección resultante satisface los objetivos que se fijaron al comienzo del ciclo de vida.

CONCLUSIONES

Desde el inicio de este proyecto, en la Biblioteca Pública del Estado en Las Palmas, que comenzó con la recopilación de sedes web para cubrir el evento de las Elecciones Generales 2016, y que aún se siguen capturando, nos encontramos en un nivel muy básico de entendimiento y aprendizaje de las herramientas ofrecidas por la BNE.

El inconveniente son los intereses jerárquicos que existen en la propia biblioteca, que impidieron en un principio la comunicación directa y transparente con los coordinadores del proyecto en la BNE y el impedimento para viajar a las reuniones que éstos celebran en Madrid.

A pesar de estas desventajas, nos basamos en el *Manual para conservadores web* que ha publicado el Área de Gestión del Depó-

sito de las Publicaciones en Línea de la BNE, además de en una reciente y cada vez más extensa bibliografía para aprender y avanzar en nuestro propósito.

Teniendo en cuenta que en la actualidad una importante parte de la producción cultural se crea en la red y que ésta crece exponencialmente, es primordial la creación de un Centro Conservador Canario de publicaciones en línea, cuyo ideal podrá derivar en una futura Biblioteca Digital Canaria, donde se recolectaría la mayor parte del patrimonio digital, garantizando su preservación a largo plazo y proporcionando acceso a los ciudadanos.

En un nivel más realista, podríamos empezar con el anteriormente definido ciclo de vida de creación de un archivo web, cuyo conjunto de prácticas se realizan y actualizan casi al mismo tiempo, evolucionando con la tecnología, y que nos hace al personal bibliotecario y a la institución crecer y evolucionar con la web que tratamos de archivar.

Este proyecto facilitaría el posicionamiento de las Bibliotecas Públicas del Estado en Canarias como cabeceras de la Red de Bibliotecas de Canarias, y facilitaría la formación de la futura Biblioteca Canaria. Esto repercutiría en proyectos beneficiosos para la sociedad, empresas, administraciones públicas y particulares.

REFERENCIAS BIBLIOGRÁFICAS

ARCO BLANCO, Ana del; ROJAS SOSA, Luna. *El Archivo de Internet: Depósito Legal de las publicaciones electrónicas tras el Real Decreto 635/2015, de 10 de julio*. Granada: Comares, 2016.

BIBLIOTECA NACIONAL (España). «Archivo de la web española» [en línea]. Disponible en: <http://www.bne.es/es/Colecciones/ArchivoWeb/index.html> [Consulta: 12-08- 2016].

BIBLIOTECA NACIONAL (España). «Archivo de la web española: Información Técnica». Disponible en línea en: <http://www.bne.es/es/Colecciones/ArchivoWeb/InfoTecnica/> [Consulta: 12-08- 2016].

BIBLIOTECA NACIONAL (España). «Depósito Legal de las Publicaciones en línea: Preguntas frecuentes». Disponible en línea en: <http://www.>

- bne.es/webdocs/Servicios/PreguntasFrecuentes/DLe_PreguntasFrecuentes.pdf [Consulta 12/08/2016].
- BIBLIOTECA NACIONAL (España). *Manual para conservadores web*. Madrid: Biblioteca Nacional de España, 2016.
- COMISIÓN EUROPEA. «Agenda digital: digitalización del patrimonio cultural de la UE para impulsar el crecimiento». Disponible en línea en: http://europa.eu/rapid/press-release_IP-11-1292_es.htm. [Consulta: 12/08/2016].
- ESPAÑA. Ministerio de Educación, Cultura y Deporte. «Real Decreto 635/2015, de 10 de julio, por el que se regula el depósito legal de las publicaciones en línea». *Boletín oficial del Estado*, 177 (Madrid, 25 de julio de 2015), pp. 62878-62885. Disponible en línea en http://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-8338 [Consulta: 04-08-2016].
- «El Estado e Internet se desinteresan del dominio '.ic'». *Canarias ahora* (10 de octubre de 2012). Disponible en línea en: http://www.eldiario.es/canariasahora/sociedad/Internet-desinteresan-dominioic_0_56695057.html [Consulta: 25/08/2016].
- EUR-LEX. «i2010: Bibliotecas digitales». Disponible en línea en: <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=uriserv:l24226i> [Consulta: 12-08-2016].
- HANNA, Kristine. «El modelo de ciclo de vida del archivado web». *Anuario A/CE de cultura digital: hacia dónde vamos, tendencias digitales en el mundo de la cultura*. Disponible en línea en: http://www.accioncultural.es/media/Default%20Files/activ/2014/Adj/Anuario_ACE_2014/7ArchivadoWeb_KHanna.pdf > [Consulta 28-08-2016].
- IFLA. *Manifiesto de las IFLA/UNESCO sobre las bibliotecas digitales*. Disponible en línea en: <http://www.ifla.org/ES/publications/manifiesto-de-las-ifla-unesco-sobre-las-bibliotecas-digitales> [Consulta: 12/08/2016].
- IPC. *Web Archiving*. Disponible en línea en: <http://www.netpreserve.org/web-archiving/overview> [Consulta: 16-08-2016].
- LLUECA, Ciro. «Archivando la web, el proyecto Padicat (Patrimonio Digital de Cataluña)». *El profesional de la información*, v. 15, n. 6 (noviembre-diciembre 2006), pp. 473-478.
- LLUECA FONOLLOSA, Ciro. «Webs siempre accesibles: las bibliotecas nacionales y los depósitos digitales nacionales». *BiD: textos universitaris de biblioteconomia i documentació*, 15 (diciembre 2005). Disponible en línea en: <http://bid.ub.edu/15lluec2.htm> [Consulta: 14-08-2016].

- PÉREZ MORILLO, Mar. «Conceptos básicos sobre archivado web». [Mensaje electrónico enviado a los conservadores web del Depósito Legal de las Publicaciones en Línea]. Disponible en línea en: <https://mail.google.com/mail/u/0/h/ijiwgjfdnp91/?&cs=wh&v=b&to=mar.perez@bne.es>”mar.perez@bne.es, 23 feb. 2016.
- PÉREZ MORILLO, Mar. «El depósito legal de las publicaciones electrónicas». *El blog de la BNE*. Disponible en línea en: <http://blog.bne.es/blog/el-deposito-legal-de-las-publicaciones-electronicas/> [Consulta: 20-08-2016].
- RED.ES. «El Ministerio de Industria, Energía y Turismo y la Biblioteca Nacional de España colaboran en la conservación del patrimonio digital español». Disponible en línea en: <http://www.red.es/redes/sala-deprensa/nota-de-prensa/el-ministerio-de-industria-energia-y-turismo-y-la-biblioteca-nacional-> [Consulta: 28/08/2016].

