

Mar Pérez Morillo, Directora de la División de Procesos y Servicios Digitales de la BNE

Entrevista realizada por CARMEN MORALES SANABRIA
Vocalía de Publicaciones de SEDIC

Mar Pérez Morillo es doctora en Latín y ha sido profesora de Filología Clásica en la Universidad de Cádiz durante seis años.

Trabaja en la Biblioteca Nacional de España desde 2003, donde ha sido responsable de la web institucional y de las redes sociales de la Biblioteca. En 2009 se hizo cargo del Archivo de la Web Española y del Depósito Legal Electrónico.

Es coordinadora del Grupo de Trabajo de Depósito Legal y Patrimonio Digital del CCB. Representa a la Biblioteca Nacional de España en el Consejo Directivo del Consorcio Internacional para la Preservación de Internet.

Desde 2018 es la Directora de la División de Procesos y Servicios Digitales de la BNE, de la que dependen el catálogo automatizado, la web institucionales, las cuentas en medios sociales, la intranet, la Biblioteca Digital Hispánica, el depósito legal electrónico y los proyectos de reutilización.



Mar Pérez Morillo

1. Trabaja en la BNE desde 2003 y desde entonces ha asumido diferentes tareas, siempre vinculadas con el entorno digital: web institucional, redes sociales, archivo de la web española, depósito legal electrónico... Una trayectoria, en definitiva, estrechamente vinculada a la preservación digital. ¿Cómo ha visto evolucionar a lo largo de estos años esta disciplina dentro y fuera de la BNE? ¿Cuáles han sido los principales hitos?

Lo primero que quería aclarar es que cuando se habla de preservación digital lo hacemos en términos muy generales, pero la preservación digital es algo muy concreto. Hablamos de preservar los contenidos de la web, pero mientras que no estén en un entorno de preservación digital y no cumplan unas características determinadas que garanticen su preservación a largo plazo no podemos hablar en puridad de preservación digital. Podríamos hablar de conservación con fines de preservación, pero hay que distinguir. A menudo la gente que se dedica a esto se queja, porque hablamos de preservación de una forma un poco frívola, para todo lo que guardamos.

Dicho esto, en la Biblioteca en concreto y en el mundo de la información en general se ha producido un gran cambio en los últimos años en materia de transformación digital, sobre todo de cara a lo que los usuarios buscan en la red y los servicios que demandan. También en cuanto a la preservación y difusión del patrimonio todo apunta a lo digital: servicios, digitalización, solicitud de reproducción de documentos, petición del carné en línea... muchos servicios bibliotecarios que ya se ofrecen en red. Y cada vez tenemos más fondo digitalizado y damos más acceso a través de la web.

El otro gran hito ha sido no solo tener fondos digitalizados, sino empezar a preservarlos con criterios más estrictos para que puedan ser consultados a largo plazo. Y también guardar la información que hay en la red. Es un paso que se ha dado en el Depósito legal, ya que también es obligación de las instituciones de la memoria guardar estos contenidos. Hemos tenido que evolucionar. Y el último paso sería la reutilización de la información. Las bibliotecas tienen que ofrecer la explotación de los datos que guardan, no solo ponerlos a disposición de la comunidad docente e investigadora o de la ciudadanía en general, sino que puedan dar lugar a otros productos.

Además, la Biblioteca tiene un patrimonio físico y digital, es como si en pocos años hubiéramos duplicado la biblioteca. Un reto, ya que no hemos duplicado nuestros recursos. Es difícil hacerlo todo con los mismos recursos e incluso con menos, porque ha habido recortes presupuestarios y falta renovar el personal que se va jubilando. Además hay que buscar una forma más integrada de gestionar las colecciones. Por ejemplo, hemos trabajado en la digitalización de nuestras colecciones, incluyendo música y audiovisuales en soporte físico, pero también tenemos música y audiovisuales en soporte físico pero en formato digital –cds y dvds–, y cuando hablamos de volcarlos a servidores para preservarlos y desvincularlos del soporte físico hablamos de proceso de digitalización, cuando un volcado de este tipo no es exactamente digitalización. Es decir, estamos empezando a cambiar nuestros procesos para gestionar nuestras colecciones de forma más fácil.

2. La BNE es una entidad de referencia en lo que a preservación digital se refiere. Nos gustaría abordar el plan de trabajo existente en este sentido: misión, objetivos, metodología y principales actuaciones.

Hablando de preservación digital en sentido estricto –número de copias de seguridad y un entorno que garantice que haya señales de alarma ante posibles deterioros–, la BNE no lleva tanto tiempo trabajando en ello. Ahora tenemos ya un volumen considerable de documentos

preservados, pero no tenemos todo. Estamos haciendo ingestas con carácter retrospectivo y nuestro objetivo es estar al día, ingestar según vamos digitalizando, pero hasta ahora no ha sido posible, porque es una realidad que llevamos mucho más tiempo digitalizando que preservando.

Los últimos proyectos de digitalización masiva, acometidos gracias al apoyo que hemos tenido de Red.es a través de varios convenios firmados en los últimos años, como es el caso del proyecto de música y audiovisuales (en cintas de vídeo VHS y Beta y en citas de casete), han sido muy importantes para nosotros. En este caso concreto se trata de soportes efímeros cuyos reproductores han quedado obsoletos. El resultado de esta digitalización ya está integrado en el entorno de preservación, es decir, todo lo que se ha ido digitalizando últimamente se ha ido integrando en este entorno.

El siguiente paso sería ingestar todo lo que tenemos nacido digital, como fruto del Depósito Legal digital, todo lo que se distribuye directamente en la Red.

Para hacernos una idea, tenemos digitalizados 1.500 terabytes y tenemos ya 870 en el entorno de preservación, poco más de la mitad. Ahora entran dos ejemplares por Depósito Legal con la nueva ley, antes eran tres, y no tiene sentido digitalizarlo todo, pero queda mucho trabajo por delante para poder decir que tenemos digitalizada la mayor parte de los fondos.

Hasta ahora se han digitalizado joyas bibliográficas, piezas únicas, obra de autores que entran en dominio público (que son muy demandadas), piezas que se prestan a exposiciones y de las que hay que hacer copias de seguridad, o si se hacen grandes exposiciones como la de Piranesi, se aprovecha para digitalizar todos estos fondos. Son las líneas principales de digitalización con las que trabajamos actualmente, y la preservación va en el mismo sentido, siguiendo a la digitalización. Nuestra intención es que vayan a la par en el medio plazo, lo que requiere de capacidad de red, almacenamiento, procedimientos y personal.

3. ¿Cómo se enfoca la preservación y depósito de las publicaciones en línea teniendo en cuenta lo efímero del patrimonio digital y la necesidad de su rápida documentación?

Aquí la exhaustividad está descartada. Es imposible guardar todo lo que está en la Red. Trabajamos en varios frentes.

Por una parte, la ley de Depósito Legal de 2011 y el Real Decreto que la desarrolla en lo que se refiere a Depósito Legal electrónico de 2015 nos dan cobertura para guardar los documentos de la web que consideremos patrimoniales, tanto de pago como libremente accesibles. Los libremente accesibles los recolecta un robot, y aquí tenemos dos vías de trabajo. En primer lugar, se trata de la recolección del dominio .es, con 1.900.000 dominios, que se hace una vez al año. Al robot se le da esta lista completa y va guardando todo lo que encuentra. Los dominios .com, .org o .edu no podemos guardarlos enteros, son dominios internacionales, pero contienen mucho patrimonio documental, objeto de Depósito Legal, con información importante sobre la historia, la cultura, la literatura y la sociedad españolas. Para esta tarea necesitamos conservadores web que seleccionen esa documentación. De estos dominios no puede hacerse una recolección masiva. La única forma es la colaboración. En el campo digital es imprescindible. Tenemos conservadores web en la Biblioteca y en las Comunidades Autónomas que van seleccionando el material a guardar: audiovisuales, fondo antiguo, Bellas Artes y Cartografía, comunidades autónomas, gastronomía, elecciones generales, política catalana... son algunas de las colecciones que hemos ido creando y que ofrecen una representatividad de lo que es la web en España. Esto, en lo referente al material libremente accesible en internet.

En cuanto a lo que no es gratuito, como plataformas de libro electrónico, música y películas españolas en la web, que no podemos vincular a un dominio, un idioma o la localización geográfica del servidor, es un trabajo para el que necesitamos la colaboración de todas las

bibliotecas patrimoniales, que tienen competencias en Depósito legal, porque son ellas las que cuentan con el personal que debe seleccionar. Hemos solicitado a diferentes plataformas que nos depositen determinados contenidos: libro electrónico de ficción, como Librandia; libro electrónico académico, como eLibro; o las normas AENOR. Tenemos en torno a 120.000 títulos y disponibles ahora mismo en nuestro gestor de objetos digitales unos 60.000.

Con todo ello vamos formando un corpus documental que ya tiene más de 460 terabytes.

4. Entrando en materia, ¿cuáles son los criterios de selección de contenidos a preservar? ¿Cómo llegan a definirlos?

En cuanto a los criterios de selección, no hay nada definido ni siquiera a nivel internacional. Primero hay que elegir de qué se quieren hacer las colecciones especiales desde una óptica patrimonial y pensando en nuestros usuarios del futuro. Por ejemplo, tenemos colecciones sobre política nacional, política catalana, cambio climático, gastronomía, patrimonio popular y estamos pensando en crear otra sobre mujer y feminismo. Son temas de interés para la opinión pública hoy.

Las bibliotecas autonómicas ponen el foco en su ámbito geográfico con colecciones que gestionan ellos y son los conservadores quienes seleccionan lo más relevante. Igual en Baleares, por ejemplo, es más relevante una colección sobre turismo mientras que en Castilla y León puede serlo sobre patrimonio histórico artístico. Es el contexto el que marca también los focos de interés a preservar.

5. ¿Podría definir el tipo de software utilizado: requerimientos, características y funcionamiento?

Para guardar la web utilizamos un software de código abierto, Netarchive Suite, creado por la Biblioteca Nacional de Dinamarca, que lo puso a disposición de la comunidad internacional. Desde entonces se ha formado un grupo de usuarios de las bibliotecas nacionales que lo utilizan para recolectar la web: Francia, Suecia y Austria, además de nosotros. En el mundo ya hay muchas bibliotecas nacionales que, desde el año 1996, guardan la web. Suecia, Australia y la Library of Congress fueron las primeras en hacerlo. Más de cuarenta países lo hacen hoy en día.

La base es un robot que se llama Heritrix, al que se le da una semilla, una url, por ejemplo www.bne.es, y que se configura con los parámetros de cuántos niveles debe descender en la navegación y con qué frecuencia debe rastrear ese sitio web. Se puede programar para que lo haga una vez a la semana o a diario, por ejemplo. También se le indica, según el lugar en el que esté la información que se quiere guardar, en qué nivel está lo que interesa de modo que capture, por ejemplo, la página de inicio más un nivel o que llegue hasta dos niveles de profundidad. Y luego muestra todo lo que encuentra tal y como si estuvieras navegando por la web aunque la información ya haya desaparecido de la web viva.

6. Háblenos del Archivo de la Web Española, funcionamiento, dimensiones, etcétera.

Comenzó en 2009 cuando la Biblioteca se dio cuenta de que tenía que guardar, como ya hacían otras bibliotecas, el patrimonio en línea. Desde 2003 existen unas directrices de la UNESCO que

urgen en este sentido y ese mismo año se crea el Consorcio Internacional para la Preservación de Internet. Pero entonces en la Biblioteca no teníamos infraestructura para abordarlo con nuestros propios medios y contactamos con Internet Archive, organización sin ánimo de lucro radicada en San Francisco, que rastrea y guarda la web del mundo entero desde 1996 y que, además, forma parte del Consorcio para la Preservación de Internet, al que nosotros nos incorporamos en 2009.

En ese ámbito de colaboración hemos recibido ayuda y soporte por parte de otras instituciones que tenían más recorrido en este campo. Entre 2009 y 2013 trabajamos así para la recolección del dominio .es, mientras que la Biblioteca mejoraba su infraestructura y buscaba la herramienta más adecuada a sus características: tamaño de la web, ámbito legal, entorno y legislación. Entre 2014 y 2015 adoptamos Netarchive Suite como herramienta de recolección y a partir de ahí comenzamos a crear recolecciones selectivas de distintos temas. La primera recolección masiva del dominio .es no se pudo hacer hasta 2016. Necesitábamos un mínimo de veinte servidores virtuales para poder hacerlo y la primera recolección masiva tardó tres meses en terminar. Ahora estamos en torno a los 45 días para recorrer y guardar el contenido de en torno a 1.900.000 dominios. Esta recolección masiva se hace una vez al año.

La otra línea de recolecciones, las selectivas, la hemos venido haciendo desde 2014 sobre eventos como, por ejemplo, la muerte de Adolfo Suárez o las elecciones. Para estas recolecciones selectivas hemos creado una red colaborativa con las Comunidades Autónomas, que tienen competencias en materia de Depósito Legal, porque no tenía sentido que un modelo tan complejo hubiera que replicarlo diecisiete veces. Lo hicimos en el ámbito del Consejo de Cooperación Bibliotecaria, dentro del Grupo de Trabajo de Depósito legal y Patrimonio Digital, ofreciéndoles a los centros de conservación autonómicos nuestro soporte, herramientas e infraestructura, para que pudieran elegir qué guardar, con qué frecuencia y con qué profundidad. Nosotros lanzamos las recolecciones desde aquí y ellos pueden ver el resultado. Esta red sigue creciendo y la mayoría de las CC.AA. trabaja así ya en esta red colaborativa del depósito legal electrónico y compartimos el repositorio de la Biblioteca.

El 20 de febrero celebramos la Jornada del décimo aniversario del Archivo de la Web. Participaron representantes de las Bibliotecas Nacionales Británica, de Dinamarca, de Países Bajos y del Archivo Web de Portugal. Si miras hacia atrás ves todo lo que hemos avanzado, aunque en el día a día siempre nos parece que vamos más lentos de lo que nos gustaría. También llevamos recopilando lo que no se recolecta libremente en la web desde 2015, cuando entró en vigor el Real Decreto que regula el Depósito Legal de las publicaciones en línea cuya consulta no podemos ofrecerla a través de Internet por motivos de derechos. A través de internet se puede saber qué contenidos están guardados pero hay que acudir a la Biblioteca Nacional o a las bibliotecas de las CC.AA. para consultarlos.

7. Explíquenos la importancia de la cooperación en el desarrollo de las labores de preservación digital desde su experiencia en el Consejo Directivo del Consorcio Internacional para la Preservación de Internet y al frente del Grupo de Trabajo de Depósito Legal y Patrimonio Digital del Consejo de Cooperación Bibliotecaria.

A nivel internacional fue fundamental integrarnos en el Consorcio para la Preservación de Internet porque se desarrollan herramientas conjuntas, se comparte conocimiento y el que tiene más experiencia asesora y ayuda a los demás. Además, en ese ámbito conocimos el proyecto de la Biblioteca Nacional de Francia, que tenía una legislación, una estructura bibliotecaria y una web muy parecida a la nuestra, y supimos que teníamos que mirarnos en ella. Asistimos a un taller en el que explicaron cómo habían acometido esta tarea, que tiene que integrarse en los

procesos bibliotecarios, igual que en su día se hizo con la digitalización, pero que se añade a todo lo que ya se viene haciendo: mantenimiento de la colección, catalogación, préstamo, restauración... En su momento, cuando se empezó con la digitalización todas las tareas relacionadas se añadían a las que ya existían y esto supuso mucho reajuste en la Institución.

Ahora además hay material en línea que hay que guardar: música, películas, videojuegos, etcétera. Supone un reto tecnológico también de almacenamiento, un cambio en los procedimientos, en la metodología y en determinar cómo se garantiza el acceso a largo plazo. Todo esto no podemos hacerlo solo con nuestros recursos. Por ejemplo, para completar nuestra colección de videojuegos estamos en contacto con las asociaciones de usuarios y con los productores de videojuegos. Y para ofrecer servicio a los investigadores estamos en contacto con plataformas digitales de Humanidades. Estos son solo ejemplos de las líneas de colaboración que debemos fomentar y ampliar para hacer frente a todos estos nuevos retos que se nos plantean.

La Biblioteca Nacional de Francia nos ha ayudado mucho, volviendo al Archivo de la Web, y el entorno de usuarios de Netarchive Suite. Cuando decidimos que era el entorno que queríamos usar, los compañeros de las otras bibliotecas participantes nos ofrecieron talleres especializados y asesoría. Trabajamos conjuntamente.

A nivel nacional hemos trabajado con Red.es, con quien hemos estado colaborando hasta finales de 2019, cuando terminó el convenio que teníamos firmado con ellos para la preservación de las colecciones digitales. Su ayuda ha sido fundamental. La infraestructura que nos ha permitido gestionar las recolecciones web con nuestros propios recursos la tenemos gracias a ellos. También gracias a Red.es llevamos a cabo el proyecto de digitalización de material audiovisual. Hay muchos hitos en este campo de la digitalización y preservación que hemos podido cubrir gracias a su colaboración.

Por otra parte, solo con el equipo que tenemos en la BNE para ello, cuatro personas en el Depósito Legal electrónico, más el personal de departamentos especiales que a tiempo parcial se dedica a la selección de este tipo de materiales, no sería suficiente, si no contáramos con la colaboración de las Comunidades Autónomas.

Por otra parte, los editores saben que las medidas que tomemos para la preservación y el acceso a estas colecciones siempre van a garantizar la protección de derechos de los autores y los intereses comerciales legítimos de las editoriales. Tenerlos como aliados en esta tarea es importante.

8. ¿Cuál es el papel de las redes sociales, páginas web y similares en las instituciones culturales como la suya?

En realidad, el primer salto digital que da la Biblioteca es su página web, desde la que muestra al mundo sus servicios, sus instalaciones y sus colecciones, y desde donde enlaza las colecciones que va digitalizando. Pero el contenido es estático y se nos quedaba pequeño. Para abrirnos a la sociedad, para transmitir que lo que estamos guardando es el patrimonio de todos, para que la sociedad lo perciba como suyo, las redes sociales son una herramienta fundamental.

No solo han permitido una mejor difusión de lo que tenemos –colecciones, contenido patrimonial, actividades, etc–, sino también la participación del ciudadano, que pide, se queja, felicita... se comunica de una forma directa con la Institución. Ese diálogo nos ha ayudado a crecer y a entender qué es lo que la sociedad pide de una institución como ésta. Lo que tenemos, que es un patrimonio de todos, debe mirar a la sociedad a la que sirve y en ese sentido las redes sociales han sido una ventana, una herramienta de comunicación importante y una vía para mostrar todo lo que guardamos y hacemos. Dentro, por ejemplo, de la Biblioteca Digital, hemos

hecho un esfuerzo en la creación de colecciones: gastronomía, joyas de la Biblioteca, literatura, teatro del Siglo de Oro... que no es más que una selección que hemos hecho nosotros pensando que al usuario puede servirle. Pero las redes sociales permiten mucho más. A propósito, por ejemplo, de cualquier efeméride podemos difundir colecciones específicas: recordatorios de comunión, postales, felicitaciones de Navidad, etiquetas de alimentos o cromos de Gallina Blanca son colecciones menos conocidas, de importancia creciente con el paso de los años, que reflejan un momento social y cultural del país. Las redes sociales nos han permitido mostrar que tenemos más que libros impresos, manuscritos o grabados. También nos han permitido abrir fuentes de investigación sobre fondos poco conocidos.

9. ¿En qué dirección cree que evolucionará el mundo digital en los próximos años y cuáles cree que serán las principales pautas para proceder a la conservación de este patrimonio?

La preservación digital es la forma de conservar este patrimonio. Para que un sitio web no desaparezca hay que recolectarlo y, a continuación, debe entrar en un entorno de preservación digital que nos garantice su conservación. La información que hay en la web es efímera y es por lo tanto contenido prioritario para ser guardado y después preservado a largo plazo. La siguiente fase sería trabajar para que sea accesible. No podemos garantizar que las webs que estamos guardando puedan consultarse dentro de treinta años, como me han preguntado ya alguna vez. Hace unos años pensábamos que la microfilmación era el mejor método de preservación, y los desarrollos tecnológicos han convertido la digitalización en el mejor sistema para preservar el contenido de nuestros fondos patrimoniales. Teniendo en cuenta que Internet Archive empezó a digitalizar la web en 1996 y ahora puedes acceder a toda esa información a través de su web, tenemos cierto margen de confianza en la sostenibilidad del sistema.

Pero habrá que estar atentos y desarrollando todo el tiempo herramientas de recuperación. En esto informáticos y expertos tecnológicos son los que deben trabajar para garantizar la recuperación y accesibilidad de este contenido. En el caso de los videojuegos, por ejemplo, se está trabajando en emuladores, para replicar la experiencia de usuario, porque no parece posible guardar miles de consolas y mantener el software y el sistema operativo necesarios para su funcionamiento a largo plazo. En cuanto a la evolución del mundo digital, en la Jornada de celebración del décimo aniversario del archivo de la web, la Secretaría de Estado de Digitalización e Inteligencia Artificial dijo que estábamos ya en una fase de postdigitalización. Debemos buscar ahora formas de explotar todo lo que hemos digitalizado y preservado, aunque tengamos que seguir digitalizando en paralelo, porque queda mucho trabajo aún. En este sentido, nosotros reorientamos la estrategia digital de la Biblioteca poniendo en marcha distintos proyectos de reutilización hace unos cuatro años y ahora tenemos dos plataformas muy activas.

Comunidad BNE es un proyecto que cuenta con la colaboración ciudadana para añadir datos que no tenemos, geolocalizar lugares que aparecen en fotos antiguas o identificar personajes que aparecen en grabados. En ella se ofrece la posibilidad de crear proyectos nuevos orientados a intereses específicos de comunidades de investigadores, por ejemplo. La otra plataforma de reutilización a la que me refería es BNEscolar, que está dirigida a la comunidad docente, con ejemplos de secuencias didácticas aplicables a diferentes niveles y también contenidos transversales en diferentes disciplinas: estereotipos que se transmiten en la publicidad, la forma en la que se cuentan determinados acontecimientos en la prensa, la diversidad cultural de la inmigración junto a temas más clásicos como la Filosofía, la Poesía o Teatro del Siglo de Oro. Se ofrecen además talleres, actividades didácticas o desafíos interactivos con aplicaciones para móvil que usan nuestro fondo digital. Se ofrecen ejemplares a partir de los cuales la comunidad docente puede elaborar sus propios materiales sobre esa plataforma.

Nosotros tenemos un corpus digitalizado y nacido digital que queremos que sea usado y explotado. Con los cerca de 460 terabytes del Archivo de la Web Española se puede hacer minería de datos utilizando herramientas de big data. En esta línea colaboramos con el Plan de Impulso de las Tecnologías del Lenguaje, de la Secretaría de Estado de Digitalización e Inteligencia Artificial, en el que se inscriben programas de procesamiento del lenguaje natural. El volumen de datos que tenemos puede servir de banco de pruebas para ver qué explotación y usos puede tener para organismos públicos y para empresas.

La digitalización, preservación y reutilización de los datos y las colecciones tienen por delante un amplio campo de aplicaciones.



Sobre la autora



CARMEN MORALES SANABRIA
Vocalía de Publicaciones de SEDIC