

## **POLÍTICA DE DESARROLLO DE COLECCIONES DEL ARCHIVO DE LA WEB ESPAÑOLA**

Biblioteca Nacional de España  
Servicio de Gestión del Depósito de Publicaciones en línea  
Versión: 2  
Fecha de actualización: 23/12/2025

## Índice

<b>1. INTRODUCCIÓN .....</b>	<b>3</b>
<b>2. MARCO LEGISLATIVO.....</b>	<b>3</b>
<b>3. ESTRATEGIA DE RECOLECCIÓN.....</b>	<b>3</b>
3.1    RECOLECCIONES MASIVAS.....	4
3.2    RECOLECCIÓN DE PUBLICACIONES SERIADAS.....	4
3.3    RECOLECCIONES SELECTIVAS.....	5
3.3.1    Temáticas .....	5
3.3.2    Evento .....	5
3.3.3    Riesgo o emergencia .....	7
<b>4. CRITERIOS DE SELECCIÓN .....</b>	<b>7</b>
4.1    CRITERIOS GENERALES DE RECOLECCIÓN .....	7
4.2    CRITERIOS ESPECÍFICOS DE RECOLECCIÓN .....	8
4.3    PUBLICACIONES EN LÍNEA RECOLECTABLES.....	8
4.4    PUBLICACIONES EN LÍNEA NO RECOLECTABLES.....	9

## 1. Introducción

Con el desarrollo tecnológico, Internet se ha convertido en el principal medio de difusión de información y es cada vez mayor la producción intelectual que se publica únicamente en línea. Este entorno digital constituye, además, el reflejo más inmediato de la época que estamos viviendo, ya que permite seguir de cerca la evolución de la investigación científica, de los movimientos sociales, de la economía, de la lengua y de la cultura. Por eso su conservación es fundamental para la investigación presente y futura.

La Unesco, en sus [Directrices para la preservación del patrimonio digital](#), publicadas en 2003, advierte del peligro de pérdida del patrimonio documental en línea. En respuesta a dicha amenaza, muchos países, conscientes de la importancia de este legado, comenzaron a adaptar sus leyes de depósito legal para incluir en él la información que se difunde a través de Internet.

En 2006 la Comisión Europea presentó su [Recomendación \(2006/585/CE\) sobre la digitalización y la accesibilidad en línea del material cultural y la conservación digital](#), en la que se instaba por primera vez a los países miembros a impulsar la digitalización de sus fondos, la creación de bibliotecas digitales y la preservación de sitios web como parte integrante de la memoria cultural europea.

Este planteamiento se reafirmó con la [Recomendación \(2011/711/UE\) de la Comisión de 27 de octubre de 2011 sobre la digitalización y accesibilidad en línea del material cultural y la conservación digital](#), en la que se anima a los Estados miembros a regular la recolección de los sitios web para reducir la carga administrativa de los productores de contenidos digitales, al tiempo que promueve la optimización de la digitalización y el uso innovador del material cultural.

## 2. Marco legislativo

A nivel nacional, la legislación sobre depósito legal contempla las recomendaciones sobre la conservación del patrimonio digital en la [Ley 23/2011, de 29 de julio, de depósito legal](#), y en la [Ley 8/2022, de 4 de mayo, por la que se modifica la Ley 23/2011, de 29 de julio, de depósito legal](#), en las que se establece que las publicaciones en línea son objeto de depósito legal, pues forman parte del patrimonio bibliográfico y, por tanto, deben ser conservadas en beneficio de las generaciones presentes y futuras. Por su parte, el [Real Decreto 635/2015, de 10 de julio, por el que se regula el depósito legal de las publicaciones en línea](#) desarrolla el procedimiento de gestión e ingreso de los recursos difundidos a través de redes electrónicas.

La preservación de este patrimonio bibliográfico corresponde de forma compartida a la Biblioteca Nacional de España y a las comunidades autónomas a través de los centros de conservación que estas designen.

## 3. Estrategia de recolección

La recolección de sitios web es la principal forma de llevar a cabo el depósito legal de las publicaciones en línea. Esta se realiza con robots rastreadores que van recorriendo las URL seleccionadas previamente y guardando el contenido, acorde a la frecuencia, profundidad y tamaño que se determinen en los parámetros de configuración. El resultado de estas recolecciones son los archivos web, donde se guarda y se puede consultar lo recolectado.

Debido al enorme volumen de publicaciones en línea existente es imposible hoy en día, con los medios tecnológicos de los que disponemos, aspirar a un depósito legal exhaustivo de estas publicaciones. Por ello, se han previsto tres formas diferentes de llevar a cabo el depósito legal de los

recursos en línea para que la preservación de este patrimonio bibliográfico sea lo más completa y eficaz posible:

- Recolectando las publicaciones directamente de la Web, utilizando un software especializado de rastreo y archivado automatizado de sitios web.
- Recibiendo las publicaciones que no puedan capturarse directamente en la Web, requiriendo al editor o productor la entrega, a través de medios electrónicos, de los recursos objeto de depósito legal en los formatos en los que estén editados.
- Creando acuerdos con repositorios y plataformas de distribución de publicaciones en línea y recursos digitales, pertenecientes a entidades públicas o privadas, que cumplan los requisitos necesarios de preservación establecidos por los centros conservadores y bajo su supervisión.

En el caso de la recolección directa de la web, y con el objetivo de guardar la mayor cantidad posible de información, la Biblioteca Nacional de España ha optado por un modelo híbrido que combina recolecciones masivas y selectivas.

Este modelo sigue la línea de las [Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo](#), y tiene como finalidad optimizar los recursos y facilitar la selección y conservación. Además, está en consonancia con otras [políticas de colecciones web internacionales](#), como es el caso de otras bibliotecas nacionales.

Estas dos categorías principales de recolección varían en el nivel de automatización, así como en el alcance y el tamaño de los archivos web resultantes.

### 3.1 Recolecciones masivas

Las recolecciones masivas capturan el mayor número de dominios posible con una profundidad relativamente pequeña en los niveles de navegación, y están vinculadas a un dominio nacional, como el dominio .es.

El objetivo es obtener una muestra representativa del panorama de la Web en un momento concreto. Para ello se sacrifica la profundidad del rastreo en beneficio de la amplitud y extensión de la cobertura del dominio.

El robot recolector, al que se le proporciona la lista de dominios registrados, los rastrea todos y archiva el contenido que encuentra a su paso, de acuerdo con una configuración predeterminada.

Las recolecciones masivas tienen por objeto rastrear un dominio completo, sin exclusiones ni selección. Esto implica que no hay una preselección por parte de los documentalistas, lo que elimina cualquier filtro, sesgo o subjetividad. Únicamente se eliminan aquellos sitios web que no existan o no puedan ser capturados en el momento que se lleva a cabo la recolección.

El Archivo de la Web Española guarda actualmente los cuatro dominios nacionales: .es, .gal, .cat, .eus. Lo hace una vez al año con la colaboración de los diferentes centros de conservación autonómicos.

### 3.2 Recolección de publicaciones seriadas

Las publicaciones seriadas en línea son uno de los contenidos más efímeros en Internet, especialmente aquellas que no tienen un respaldo institucional. Con el objetivo de conservarlas se realiza un guardado al año de todas las publicaciones seriadas registradas con ISSN en España y en libre acceso en Internet. Esta tarea se ve facilitada al ser la Biblioteca Nacional de España el Centro

Nacional del ISSN, por lo que su catálogo incorpora todas las nuevas publicaciones que reciben este número internacional.

A diferencia de las recolecciones masivas por dominios, en este caso se busca una mayor profundidad de rastreo para lograr una captura tan completa como sea posible de los ejemplares digitales de las revistas y se lleva a cabo una mayor labor de control de calidad.

### 3.3 Recolecciones selectivas

Las recolecciones selectivas complementan a las masivas, ya que recogen con mayor profundidad y frecuencia una muestra más reducida de sitios web seleccionados por su relevancia para la historia, la sociedad y la cultura.

Esta forma de recolección está respaldada en las [Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo](#), donde se señala la selección de sitios web como estrategia para recopilar contenidos digitales.

La selección se realiza en colaboración con los centros de conservación de las comunidades autónomas, también depositarios del depósito legal de publicaciones en línea. En el marco de colaboración con el Consejo de Cooperación Bibliotecaria, los centros de conservación seleccionan y realizan el control de calidad de los sitios web representativos para sus respectivas comunidades autónomas.

Las recolecciones selectivas que se realizan son de tres tipos:

#### 3.3.1 Temáticas

Capturan una selección de recursos sobre un tema de forma periódica. El propio Servicio de Depósito Publicaciones en Línea de la Biblioteca Nacional de España y los centros de conservación de las comunidades autónomas mantienen sus colecciones temáticas con los recursos en línea que consideren necesario conservar como parte del depósito legal, siempre que cumplan los criterios de selección.

Como complemento de las colecciones impresas, se forman colecciones temáticas basadas en los fondos de las salas especiales de la BNE y con una selección creada por los propios especialistas de estas salas: Bellas Artes y Cartografía; Música y Audiovisuales; Manuscritos, Incunables y Raros; Revistas electrónicas.

#### 3.3.2 Evento

Sobre acontecimientos de especial relevancia para la sociedad española y con una duración determinada, que reúnen los siguientes requisitos:

- Especial relevancia por su valor cultural, científico, artístico, social y político para la sociedad española de hoy y del futuro.
- Trascendencia para la investigación futura sobre la historia, la sociedad y la cultura españolas.
- Gran repercusión pública e impacto social.

El interés de estos eventos debe ser contemplado siempre desde una perspectiva de futuro. Se debe valorar, pues, si son temas que se podrán considerar trascendentales y de repercusión nacional o internacional pasados unos 20-50 años.

Los eventos sobre los que la Biblioteca ha establecido realizar recolecciones, siempre que cumplan los criterios de selección, son los siguientes:

- Eventos políticos: procesos electorales europeos, nacionales, autonómicos y locales; cambios en la Jefatura del Estado; cambios legales de gran trascendencia...
- Fallecimientos de personalidades destacadas.
- Catástrofes: accidentes de gran magnitud, atentados terroristas, terremotos, desastres ecológicos, etc.
- Tragedias humanitarias: epidemias, crisis sanitarias...
- Eventos culturales: no sólo de importancia general (Premio Cervantes, Princesa de Asturias...), sino también de impacto social o mediático.
- Eventos deportivos: Juegos Olímpicos, campeonatos mundiales, competiciones de alto impacto mediático celebrados en España, etc.

Determinar el valor a largo plazo e implicaciones futuras de un acontecimiento es de suma importancia, aunque no siempre resulta fácil de prever. No existen reglas definitivas que permitan anticipar con certeza el impacto de un hecho concreto. Sin embargo, se pueden considerar algunos criterios orientativos que facilitan esta valoración, como la magnitud del evento (número de sitios web creados y cobertura en medios de comunicación), la cobertura en otras colecciones temáticas, la repercusión en redes sociales, así como su influencia política y social a lo largo del tiempo. Su aplicación tiene una parte subjetiva, por lo que se recomienda que la valoración sea hecha por un equipo diverso de personas.

Para ampliar y hacer más objetiva esta visión es fundamental contar con la colaboración de las comunidades autónomas. Además, se está trabajando para habilitar otras vías de colaboración, por ejemplo, con instituciones especializadas en distintas áreas del conocimiento, y con los ciudadanos, de manera que puedan sugerir sitios web para conservar como parte del patrimonio digital español.

La decisión de crear una colección exclusiva para un evento corresponde al Servicio de Gestión del Depósito de Publicaciones en Línea de la BNE, basada en la trascendencia y repercusión del acontecimiento a nivel nacional, así como en su impacto internacional. En el caso de eventos con un notable impacto social y mediático que afecten a una o varias comunidades autónomas, los centros de conservación de estas CC. AA. podrán proponer la creación de colecciones de evento y se tomarán en cuenta sus opiniones a la hora de decidir si se crea una colección específica o si el evento se integra en sus colecciones autonómicas.

La Biblioteca Nacional de España participa en recolecciones colaborativas internacionales organizadas por el IIPC (International Internet Preservation Consortium) con motivo de acontecimientos de interés transnacional, por ejemplo, la pandemia del Coronavirus, los Juegos Olímpicos de Tokio 2020, el ataque contra Charlie Hebdo o la guerra de Ucrania...

La cobertura de eventos no siempre conllevará la creación de una colección, sino que, en algunos casos, estos serán cubiertos como parte de una colección ya preexistente de temática afín; sirva de ejemplo la jura de la Constitución de la Princesa Leonor, cubierta dentro de la colección Política Nacional, o el fallecimiento del dibujante Francisco Ibáñez, que forma parte de la colección Cómics.

Un caso especial son las elecciones, que siempre tendrán una colección propia debido a la gran información en sí misma que generan y su importante repercusión política.

### 3.3.3 Riesgo o emergencia

Las recolecciones de riesgo se realizan en el caso de cierre de sitios web que impliquen pérdida de documentación y publicaciones en línea. En estos casos se procede a realizar un guardado exhaustivo que permita recoger la mayor cantidad posible de información del sitio que va a desaparecer, puesto que, tras su eliminación esta información se perderá irremediablemente.

El servicio de publicaciones en línea sigue la actualidad y rastrea la Web buscando sitios que anuncien su desaparición, pero también es importante la colaboración ciudadana, pudiendo informar de la desaparición de un sitio web a través del correo: [archivoweb@bne.es](mailto:archivoweb@bne.es). Algunos ejemplos son el guardado de Wikispaces o los foros de Yahoo.

Son también importantes para su conservación los sitios web que vayan a ser profundamente reformados, ya que las versiones anteriores proporcionan una gran información de una época determinada en la historia de Internet.

En el caso de recolecciones de emergencia, la finalidad es captar toda la información posible de un evento que sucede en tiempo real y cuya información se produce en grandes cantidades en muy poco tiempo. Esto implica la reacción rápida del equipo de conservadores web para salvar toda la información posible y que sirva como una fotografía de los sucesos acontecidos. Un ejemplo de esto son los atentados de Cataluña de 2017.

## 4. Criterios de selección

La inmensa cantidad de información publicada en Internet hace imposible que se conserve todo, por eso es necesario establecer unos criterios para determinar qué contenidos se consideran prioritarios para ser preservados en las recolecciones selectivas.

A nivel internacional no existen aún criterios de selección normalizados para este propósito, aunque las instituciones patrimoniales de todo el mundo que guardan la Web trabajan intensamente en este campo. Por el momento, las organizaciones establecen sus criterios de selección en función de las políticas institucionales, los recursos de personal, las posibilidades técnicas, la capacidad de procesamiento informático y la de almacenamiento. Además, hay en algunos casos limitaciones derivadas de la legislación.

Las [Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo](#) proponen algunos criterios de selección, entre los que se encuentran el área temática, el tipo o formato del contenido, los creadores de una región o movimiento determinados, y la función que desempeñan los sitios web.

Teniendo en cuenta lo anterior, 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.

### 4.1 Criterios generales de recolección

Los criterios generales de selección están basados en el artículo 3 del Real Decreto 635/2015, regulador el depósito legal de publicaciones en línea, según el cual los sitios web objeto de depósito legal son aquellos que:

- contengan patrimonio bibliográfico, sonoro, visual, audiovisual o digital de las culturas de España;

- estén bajo el dominio .es y subdominios asociados, así como otros dominios del territorio nacional;
- estén albergados en otros dominios (.com, .net, .edu, .org, etc.), pero contengan patrimonio documental español;
- estén en cualquiera de las lenguas oficiales del Estado;
- estén en cualquier formato, incluyendo las publicaciones en ellos contenidas;
- sean tanto de acceso libre como restringido.

## 4.2 Criterios específicos de recolección

Como criterios específicos de selección para realizar una recolección selectiva se han establecido los siguientes:

- sitios web que se consideren de especial relevancia patrimonial, para la investigación y el conocimiento de nuestra sociedad en el futuro;
- sitios web que contengan información relacionada directamente con el tema de la recolección;
- se priorizan los contenidos en línea que tengan riesgo de pérdida por desaparición de la página en la que estén albergados;
- se prioriza la inclusión de recursos que no tengan una versión impresa y que por tanto se perderían si no se guarda la versión en línea;
- para las recolecciones de evento se priorizan los sitios web con actualización más reciente o que se actualicen con más frecuencia;
- se guardan sitios web que cubran todos los puntos de vista posibles de una temática concreta, sin juzgar las ideas expuestas en los sitios seleccionados;
- no se juzga la veracidad de la información.

## 4.3 Publicaciones en línea recolectables

Existen varias categorías de recursos web que es recomendable incluir en las selecciones de sitios web a recolectar para que la muestra documental sea lo más representativa posible:

- Medios de comunicación social: cabeceras de prensa, agencias de noticias, cadenas de radio y televisión.
- Organismos administrativos: ministerios, comunidades autónomas, ayuntamientos.
- Instituciones políticas: partidos políticos.
- Instituciones culturales: museos, archivos, bibliotecas, colegios, universidades, centros de investigación.
- Instituciones científicas.
- Instituciones sanitarias.
- Instituciones deportivas.
- Sitios web centrados en patrimonio natural y artístico.
- Actos culturales, congresos, asambleas, jornadas...
- Empresas privadas.
- Asociaciones: profesionales, ONG, etc.
- Blogs y páginas web de personas relevantes relacionadas con el tema de la recolección.
- Redes sociales: Twitter (actual X), Facebook.
- Wikis: Wikipedia.
- Grabaciones en vídeo: YouTube.

Estos grandes grupos son en general los tipos de sitios web que puede cubrir una colección o evento, pero según la colección pueden aparecer sitios web no contemplados en esta lista y que sea interesante añadir para una colección o evento determinado. La decisión de su inclusión queda a juicio y valoración del conservador web, basado en su importancia para la interpretación del hecho o la colección para el futuro.

#### 4.4 Publicaciones en línea no recolectables

Según el Real Decreto 635/2015, quedan excluidos de las recolecciones (art. 4):

- Los correos y la correspondencia privada.
- Los contenidos que estén albergados únicamente en una red privada.
- Los datos personales a los que sólo tiene acceso un grupo restringido de personas.

De conformidad con lo dispuesto en los artículos 6 y 7 del Real Decreto 635/2015, de 10 de julio, la BIBLIOTECA NACIONAL DE ESPAÑA, O.A., ejerce su función de captura y depósito de las publicaciones en línea que hayan sido objeto de comunicación pública y los sitios web accesibles a través de redes de comunicaciones. Esta captura y depósito se realiza sin alteración de los contenidos, con el propósito de garantizar su integridad y trazabilidad histórica. En consecuencia, la BNE no se responsabiliza de aquellos contenidos que, formando parte de la captura y depósito realizado, sean contrarios a la ley, la moral o el orden público, siendo responsables de aquellos los titulares de dichas comunicaciones.