



Los archivos web y el patrimonio nacido digital: el caso de la Biblioteca Nacional de España

José Carlos Cerdán Medina

Jefe del Servicio de Gestión de Depósito de Publicaciones en Línea de la Biblioteca Nacional de España, España

josec.cerdan@bne.es

Resumen: La Biblioteca Nacional de España creó en 2009 el Archivo de la Web Española con la finalidad de comenzar a guardar Internet y las publicaciones que en ella se albergan y preservarlas para futuras generaciones. Los contenidos difundidos en Internet se caracterizan por producirse en cantidades inmensas diariamente, mucho mayores que el impreso; también son contenidos altamente efímeros y volátiles; y mucho de ellos, se crean de una manera constante y su desaparición no deja rastro. La colaboración a nivel nacional e internacional es un elemento clave en el desarrollo del archivo web. Un proyecto joven, que, con 15 años, ya supera los 2 Petabytes de información guardados de Internet. Al archivo de la web se une el ingreso de publicaciones nacidas digitales que nos son recolectables automáticamente por encontrarse en venta o en acceso restringido. En conclusión, los archivos de la web se sitúan como una solución tecnológica de primer orden en la que cimentar el desafío que suponen la proliferación de contenidos nacidos digitales en Internet, que en muchos casos nunca llegarán a ser impresos, y para asegurar la conservación del patrimonio y la memoria entre generaciones.

Palabras clave: Patrimonio nacido digital, Archivos web, Bibliotecas nacionales, Archivo de la Web Española, Biblioteca Nacional de España



El desafío de los contenidos nacidos digitales

A mediados de los 90, surgió el Internet comercial y, aunque en aquel momento no éramos conscientes, algo cambió. Se democratizó la comunicación y cualquiera podía difundir información de manera global. Después llegaron las redes sociales, y esto, se convirtió en algo cotidiano e instantáneo.

En este desarrollo se creó una gran cantidad de documentos “nacidos digitales”, unos basados en el nuevo lenguaje de la web, la hipertextualidad, y otros, más tradicionales imitando el impreso en formatos como el PDF. De hecho, cuando nació la imprenta, los primeros impresos, los incunables, buscaron parecerse a los manuscritos de “gran tradición” y mayor prestigio. La evolución de Internet no fue demasiado distinta.

A lo largo de los años Internet alcanzó su madurez, aunque de manera mucho más rápida que cualquier otra tecnología, y se convirtió en la imprenta personal y portátil más potente de todos los tiempos. Cualquiera puede crear contenido y difundirlo rápidamente con unos conocimientos básicos de informática, y con ello, la autoedición alcanzó cuotas nunca antes conocidas, aupada especialmente, por la facilidad de distribución que daba Internet.

En este escenario que avanza tan rápidamente, se encuentran las bibliotecas nacionales, las cuales deben cumplir con el deber de preservar el patrimonio y la memoria, sin importar el soporte en que se encuentre, cumpliendo su función de ser garantes de su conservación para las generaciones futuras.

Pero independientemente de que hablemos de sitios web o libros electrónicos, ¿a qué desafíos se enfrenta una biblioteca, a diferencia de los documentos impresos? Podemos resumirlos en cuatro principalmente:

- Una producción masiva, nunca antes alcanzada, con un carácter mucho más efímero que hace imposible su conservación total.
- Es necesario diferentes herramientas tecnológicas para su guardado e interpretación.
- La seguridad informática, en algunos casos, son productos comerciales de alto valor económico que no pueden ser almacenados o difundidos sin las medidas de seguridad adecuadas.
- La preservación. Un libro impreso tiene sus complicaciones específicas de conservación, pero un libro digital le debemos sumar los problemas de almacenamiento (migración) y los problemas de lectura de formatos (transformación o emulación)

En este escenario, las bibliotecas nacionales deben emprender verdaderas carreras contrarreloj para conservar lo máximo posible de una tecnología reciente, crucial para la interpretación de nuestra sociedad, y que muy difícilmente podrá cubrirse con el material impreso. Sirva de ejemplo el proyecto de arqueología digital de la Biblioteca Real de Holanda con su colección de sitios web entre 1994 y 2001, conocido como XS4ALL e incluido dentro del listado



de patrimonio mundial de la UNESCO. Algo increíble para una tecnología que no supera los 30 años de antigüedad y que da muestras de su importancia futura.

Qué es un archivo de la web

En este escenario, era necesario desarrollar herramientas que permitieran guardar sitios web y reproducirlos tal como eran, con esa finalidad nacen los archivos web.

El International Internet Preservation Consortium (2024), el organismo internacional más importante en la conservación de la web, define el archivado web como: “el proceso de recopilar partes de la World Wide Web, preservar las colecciones en un formato de archivo y, a continuación, servir los archivos para su acceso y uso”

Los archivos de la web permiten guardar, no sólo los sitios web, sino también las publicaciones y documentos que en ellos se albergan. Esto los convierten en el mejor aliado de las bibliotecas para guardar el patrimonio documental nacido digital, lo que incluye no sólo los sitios web, sino todo lo que se produce en Internet en diferentes formas y se encuentre en acceso libre, como son los folletos y libros digitales, las revistas, las grabaciones sonoras, los vídeos e incluso las redes sociales.

Por tanto, los archivos web se convierten en la base de la preservación del patrimonio nacido digital y un aliado para su recolección y almacenaje.

Contexto internacional

En 1996 nace, el que es sin duda, el padre de todos los archivos web Internet Archive. Su creador Brewster Khale lo creó con la vocación de guardar Internet y de ofrecer todo su conocimiento de manera universal.

A pesar de que comenzara todo a partir de una organización sin ánimo de lucro, considerada por su propio creador como “la biblioteca de la gente” (Khale, 2021), a nivel mundial, esta tarea recayó en las instituciones que hasta este momento habían preservado el patrimonio y la memoria de manera generalizada, las bibliotecas nacionales. La primera biblioteca en tener archivo web fue la Biblioteca Nacional de Suecia (Arvidson y Lettenström, 1998), que también formó parte de la Nordic Web en 1997, junto a: Islandia, Noruega, Dinamarca y Finlandia. En torno al año 2000, será cuando otras bibliotecas nacionales crearan sus archivos web, como Australia, Reino Unido, Francia...

Crucial será el año 2003, cuando la UNESCO tome conciencia del desafío que supone guardar Internet y la gran cantidad de patrimonio que estaba desapareciendo. Las directrices para la preservación del patrimonio digital (Webb, C., 2003), supusieron una voz de alarma; ese mismo año, se creó también el International Internet Preservation Consortium (IIPC), ambas instituciones pusieron sobre la mesa, que los documentos nacidos digitalmente, páginas web incluidas, debían ser preservadas a largo plazo, como se venía haciendo con las colecciones impresas. Proyectos más recientes de la UNESCO como la Plataforma PERSIST, con sus



Directrices para la selección del patrimonio digital para su conservación a largo plazo (Parent et al., 2021) han supuesto un análisis sobre las decisiones de selección y el desafío únicos de su conservación.

La Unión Europea en 2006, consciente del valor del patrimonio nacido digital, en su recomendación de la Comisión sobre la digitalización y accesibilidad en línea del material cultural y la conservación digital, habla de la importancia de que los estados miembros tengan en cuenta al resto de estados en sus políticas de ingreso de documentos nacidos digitales y desarrollen legislación para el archivado web (Comisión Europea, 2006). En 2011 se vuelve a hacer hincapié en la importancia del ingreso de los documentos nacidos digitales y el guardado de sitios web, con especial énfasis en su conservación a largo plazo (Comisión Europea, 2011). En este ámbito se produce la creación del Archivo de la Web Española en 2009 con el fin de conservar Internet y las publicaciones que en él se albergan, como parte del patrimonio cultural y científico.

Características especiales de los contenidos en Internet

Los contenidos en Internet son especiales en sí mismos, no son comparables a cualquier otro tipo de documentos digitales, por las siguientes razones:

- Cantidad: una producción enorme y continua. Esto hace que su guardado exhaustivo o total quede descartados y se tienda a la selección por diversas estrategias y técnicas.
- Duración: son muy efímeros y tienden a desaparecer con facilidad.
- Actualización: los sitios web cambian continuamente y su guardado debe producirse de manera periódica.
- Perdida: su desaparición es total e irrecuperable, si no es a través de un archivo web

Estas características, hace que los trabajos en el guardado de sitios web necesiten estrategias y formas de trabajo muy diversas a los de otros contenidos nacidos digitales, como, por ejemplo, los libros digitales.

Utilidad de los archivos de la web

- El archivo de la web permite visitar los sitios web archivados y navegar en ellos, tal y como se podían encontrar en la web viva. Sin querer ser minucioso, las utilidades más habituales de un archivo web son las siguientes
- Historia de Internet. El poder visitar un sitio web en diversos momentos de su historia permite ver la evolución de las tendencias y modas en este campo.
- Información sobre eventos y temáticas. Fuente de información primaria de acontecimientos importantes en un país, con información crucial para su interpretación, que nunca llegará a estar impresa.
- Huella en Internet. Recuerdo de un sitio web desaparecido con información importante para un usuario o un público determinado.



- Banco de publicaciones digitales desaparecidas. Guardado de documentación, especialmente oficial: catálogos, folletos, informes... producidos en un momento determinado y que tienden a desaparecer. También importante en las publicaciones de autores que difunden su obra de manera libre en Internet.
- Banco de datos para investigación e inteligencia artificial. El Archivo de la web almacena gran cantidad de contenido de un lenguaje actual y moderno, muypreciado para investigadores relacionados con el lenguaje o el entrenamiento de la Inteligencia Artificial (Pastrana y Cerdán 2021)

Legislación española en el ámbito del depósito digital

La legislación para la salvaguardia del patrimonio nacido digital es bastante reciente. En España, bien es cierto, que el decreto de 1957, que inicia el depósito legal moderno, fue consciente de los grandes cambios tecnológicos que estaban por venir y su redacción abierta, permitió guardar contenidos nacidos digitales de forma temprana. No será hasta la ley de 2011 cuando se nombre por primera vez los sitios web y las publicaciones en línea como soporte electrónico no tangible. La ley quedaba pendiente en este ámbito, de un desarrollo posterior por un decreto específico.

Cuatro años más tarde, en 2015, se desarrolló el primer real decreto centrado en el depósito legal de publicaciones en línea. En él, se profundiza y delimitan varios aspectos fundamentales para la conservación del patrimonio nacido digital:

- **Ámbito de aplicación:** se especifica que puede ser guardado y que no.
- **Alcance:** Se descarta la exhaustividad, debido a la imposibilidad de guardar todo este patrimonio.
- **Cambio de paradigma:** las publicaciones no deberán ser ingresadas por los editores, ni recibir un número de depósito legal para ser publicadas. Serán la BNE y los centros de conservación los que seleccionen y se dirijan a editores y productores para el ingreso.
- **Modelo:** La vía preferente de guardado es el archivado web. Con la obligación de los editores o productores de permitir el guardado automático.
- **Obligaciones legales:** los editores o productores deberán dejarse guardar o entregar las publicaciones en caso de que no sea posible hacerlo de manera automática.
- **Colaboración:** Posibilita los convenios de conservación entre instituciones y centros de conservación para no duplicar esfuerzos, asegurando a su vez la preservación de los fondos a largo plazo.

Finalmente, en 2022 se modificó la ley de 2011, cuyo mayor cambio es la de exigir la copia previa a la impresión a la vez que se ingresa el impreso, pero en el terreno puramente digital, no hizo más que reforzar el papel del Real Decreto de 2015 y aclarar algunos conceptos de lo que se considera un sitio web.



El Archivo de la Web Española

Desde la Creación del Archivo de la Web Española en 2009, se decidió llevar a cabo una estrategia híbrida (Llueca, 2005) en al que se combinaran las recolecciones masivas de dominio y las selectivas sobre colecciones y eventos.

Las recolecciones de dominio suponen un “muestreo representativo” (Parent et al., 2021) de la web, normalmente, de un nombre de dominio nacional con un límite tamaño por dominio único, por lo que no deja de ser superficial. En el caso español, llevamos a cabo cuatro guardados de dominio, siendo el de mayor tamaño el del dominio .es, con más de 2.000.000 de dominios en 2023, pero también se llevan a cabo aquellos dominios representativos de las lenguas cooficiales: como el .gal para el gallego, el .cat para el catalán o el .eus para el euskera.

Este tipo de recolecciones dan una visión general de un dominio en un momento determinado y, llevadas a cabo de manera anual, permiten rastrear el estado de la web en ese dominio a lo largo del tiempo. Al guardarse todo tipo de sitios web supone la forma de guardado más objetiva y completa.

Mientras que el guardado selectivo viene a complementar las masivas. Esto se ejecuta por personal especializado, normalmente proveniente del mundo bibliotecario o de la documentación, conocidos como conservadores web, que seleccionan, en base a unos criterios, un grupo de sitios web que puedan ser por su interés cultural, sociológico, histórico o artístico, representativos para cubrir todos los puntos de vista de una temática determinada. Los archivos web suelen crear políticas de creación de colecciones que documentan los criterios para la formación de estas. En el caso de la BNE, aunque existía, desde hace muchos años de manera interna, se hizo pública en 2023 (Biblioteca Nacional de España, 2023).

Las recolecciones selectivas, en el caso de la BNE, se dividen en tres tipos:

- Temáticas: responden a dos tipologías principales, aquellas que se ocupan de una tipología documental como es el caso de la prensa, ya sea nacional o local, y las que se ocupan de un campo temático más o menos específico, como es el caso de Política Nacional, Bellas Artes y Cartografía, o más específicas como Feminismo, Videojuegos o Cómics. Un apartado especial, supondrían aquellas colecciones relacionadas con las comunidades autónomas (diferentes entidades territoriales) que tienen una temática general, delimitadas por una circunscripción geográfica o política.
- Eventos: las colecciones de eventos se caracterizan por tener un principio y fin temporal definidos, y tratan de ser una fotografía que represente lo ocurrido en periodo específico sobre un hecho determinado. Hay principalmente dos tipos; por un lado, los relacionados con elecciones: europeas, nacionales, autonómicas y locales que se vienen cubriendo sistemáticamente desde 2014. Por el otro, los acontecimientos de importante impacto para la historia de España, como es el caso de la Coronación de Felipe VI o la Pandemia del Coronavirus.



- Riesgo: creada, principalmente, para evitar la pérdida de sitios web de manera definitiva, que pudieran tener interés para el futuro. Los dos casos más importantes hasta la fecha son: Wikispaces, servicio de alojamiento web gratuito, donde cualquier persona podía crear su wiki sobre diversos temas, desaparecida en 2018, o los foros Yahoo, donde se preguntaban y respondían gran cantidad de cuestiones entre usuarios, que dejó de funcionar en 2019.

Las recolecciones selectivas buscan la cobertura de un tema desde todos los puntos de vistas y con todo tipo de fuentes. La variedad de tipologías de sitios web guardadas es enorme y muy variada: medios de comunicación como prensa, radio y televisión, sitios web oficiales, instituciones políticas, culturales, educativas, científicas, sanitarias, deportivas, sitios web de patrimonio histórico, artístico y cultural, asociaciones profesionales, ONGs, redes sociales, vídeos, podcast...

No existen unos criterios de selección internacionales, más allá de las sugerencias por la plataforma PERSIST en su documento de 2021 (Parent et al.), por lo que tienen un cierto grado de subjetividad, basada en los siguientes principios, aplicados por los conservadores web (Biblioteca Nacional de España, 2023):

- 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.

Las diferencias de temáticas y tipologías documentales de cada colección hacen difícil el crear criterios muy específicos. Por supuesto, es imposible saber, a ciencia cierta, que sitios web serán claves para el desarrollo una investigación en el futuro, pero es indudable, que la cobertura de un tema determinado de manera extensa, junto a las recolecciones masivas, dan un campo de trabajo que permite recrear un momento determinado sobre una temática o un hecho específico, de ahí, que en muchas ocasiones se comparen los archivos web con máquinas del tiempo.



Colaboración

Una de las grandes claves del Archivo de la Web Española es la colaboración. Hubiera sido imposible la cobertura de tantas temáticas, muchas de ellas de manera muy específica, sin la colaboración con multitud de instituciones públicas y privadas.

La colaboración presta un apoyo insustituible en dos campos principalmente:

- Conocimiento: colaborar con instituciones que conozcan ciertas temáticas, ayuda a crear una mejor selección y una mayor relevancia de los contenidos guardados.
- Personal: el apoyo con recursos humanos, aunque sea a tiempo parcial, permite ampliar la variedad de colecciones y permite mejorar los resultados en la inversión en un proyecto de estas características.

Las colaboraciones con instituciones externas se materializan por tres vías:

- Consejo de Cooperación Bibliotecaria: es el órgano que canaliza la colaboración bibliotecaria en España. Nos permite estar en contacto con los diferentes centros de conservación y contar con más de 40 conservadores web con formación bibliotecaria, que apoyan, principalmente, en la creación de las colecciones de sus regiones y ciertos eventos como por ejemplos las elecciones nacionales, autonómicas y locales.
- Rebiun: la Red de Bibliotecas Universitarias Española participa con 10 voluntarios de perfil bibliotecario, que permite abordar colecciones relacionadas con el mundo universitario y sobre temática científica, infrarrepresentadas habitualmente en los archivos web, más centrados habitualmente en las ciencias sociales y humanidades
- Fundación Sancho el Sabio: uno de los centros de documentación más importantes del mundo sobre cultura vasca, nos apoya haciendo una selección y control de calidad muy cuidada, sobre sitios web relacionados con el País Vasco.

La colaboración también debe ser interna, en la propia Biblioteca Nacional hay perfiles bibliotecarios altamente especializados en ciertas materias que pueden ayudar a crear buenas colecciones. En el caso de la BNE, hay cuatro principales:

- Bellas Artes y Cartografía
- Música y Audiovisuales
- Manuscritos, Incunables y Raros
- Documentación bibliotecaria

Estas colecciones sirven para completar el amplio acervo que entra en la biblioteca a través del “depósito legal tradicional” con contenidos producidos en Internet y cuyo valor patrimonial es muy importante para estas temáticas.

En todos estos casos, la BNE coordina el proyecto, poniendo una infraestructura tecnológica y de almacenamientos, proveyendo de formación especializada en este campo y con la creación de una red de puntos consultas distribuidos por toda la geografía española. Mientras



que los colaboradores participan con recursos humanos que se encargan principalmente de la selección y control de calidad de los contenidos guardados.

Sin duda, los frutos de la colaboración han sido muy importantes hasta la fecha, sin descartar ampliar la colaboración a nuevas instituciones, con un modelo que ha demostrado ser exitoso.

Contenidos no recolectables automáticamente

El Archivo de la Web Española es la base de la conservación de contenidos digitales, pero muchos contenidos, ya sea por encontrarse en acceso restringido o porque su guardado no sea posible técnicamente.

La estrategia llevada a cabo desde 2016, ha sido la de colaborar con grandes distribuidoras de contenidos digitales que pudieran ingresar gran parte de este tipo de documentos y sus metadatos de manera masiva, descargando a sus editores de la tarea de ingresar sus documentos de manera individual y facilitando a la BNE su ingreso y tratamiento masivo.

Actualmente, por esta metodología, se han ingresado gran parte de la producción del libro digital y los audiolibros españoles en diferentes tipologías, que incluyen: el libro de ficción, las publicaciones académicas y de organismos oficiales o las normas UNE, traducción de las famosas normas ISO a su versión española, creadas por los diferentes comités de normalización. Se han superado los 200.000 ingresos y se ha creado un flujo de transformación masiva que permite la integración de gran cantidad de publicaciones en el catálogo siendo la única intervención humana, la normalización de autoridades.

Se está empezando a trabajar para el ingreso de nuevas tipologías documentales como la música digital o el podcast, los cuales nos hace enfrentarnos a nuevos desafíos desconocidos hasta el momento, dada las grandes diferencias en su producción y distribución que estos tienen con el libro digital.

Patrimonio y memoria

La conservación del patrimonio nacido digital, según lo abordado anteriormente, tiene unas peculiaridades que lo hace muy distinto al tradicional depósito legal físico, que hace que no se puedan aplicar las mismas medidas o flujos de trabajo que se desarrollaron para este. Pero, a la vez, es innegable que los documentos nacidos digitales albergan un importante valor como patrimonio cultural y documental, que debe ser guardado y conservado.

El archivo de la web como capsula del tiempo de los contenidos producidos en Internet, es una herramienta valiosa para la interpretación de nuestra sociedad y la historia de Internet; un legado que cobrará valor con los años, ya que muchos de estos sitios web y documentos habrán desaparecido y sólo se podrán acceder a aquellos conservados como parte del patrimonio nacido digital de cada país. En los archivos web, no sólo se guardarán los objetos



digitales producidos, sino también el entorno en el que fueron creados, ayudando a entender los fines que tuvieron y el tiempo que perduraron.

A la vez, los contenidos digitales son memoria de los eventos y vicisitudes de un país a lo largo de su historia y será el papel de los investigadores bucear en ellos para comprender mejor los acontecimientos del pasado y las connotaciones de una sociedad, cada día, más marcada por la información producida en Internet, y que será cada vez más difícil de entender, sin una representación de todo lo producido en este entorno.

Por tanto, patrimonio y memoria hacen que el depósito legal de las publicaciones en línea sea un desafío crucial en el ámbito de las bibliotecas nacionales, que debe de ser abordado para evitar su destrucción o pérdida definitiva para las generaciones futuras, como parte de nuestro patrimonio.

Bibliografía

Arvidson, A., y Lettenström, F. (1998). The Kulturarw Project—The Swedish Royal Web Archive. *The Electronic Library*, 16(2), 105-108. <https://www.emerald.com/insight/content/doi/10.1108/eb045623/full/pdf> [Consulta: 14/06/2024]

Biblioteca Nacional de España (2023). Política de desarrollo de colecciones del Archivo de la Web Española (15/01/2024). https://www.bne.es/sites/default/files/repositorio-archivos/politica_desarrollo_colecciones_awe.pdf [Consulta: 14/06/2024]

Comisión Europea (2006). Recomendación de la Comisión, de 24 de agosto de 2006, sobre la digitalización y la accesibilidad en línea del material cultural y la conservación digital. Diario Oficial n L, 236, 0028-0030. <https://www.boe.es/doue/2006/236/L00028-00030.pdf> [Consulta: 14/06/2024]

Comisión Europea. (2011). Recomendación de la Comisión, de 27 de octubre de 2011, sobre la digitalización y accesibilidad en línea del material cultural y la conservación digital. Diario Oficial L 283 de 29.10.2011, p. 39–45, de 29.10.2011. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32011H0711> [Consulta: 14/06/2024]

International Internet Preservation Consortium (15/01/2024). WEB ARCHIVING. IIPC netpreserve.org | International Internet Preservation Consortium. <https://netpreserve.org/web-archiving/>. [Consulta: 14/06/2024]

Kahle, B. (2021). Turns Out It's Not the Technology, It's the People. Internet Archive Blogs. <https://web.archive.org/web/20211121161831/https://blog.archive.org/2021/10/22/turns-out-its-not-the-technology-its-the-people/> [Consulta: 14/06/2024]

Ley 23/2011, de 29 de julio, de depósito legal. *Boletín Oficial del Estado*, 182, de 30 de julio de 2011. <https://www.boe.es/buscar/doc.php?id=BOE-A-2011-13114> [Consulta: 14/06/2024]



Ley 8/2022, de 4 de mayo, por la que se modifica la Ley 23/2011, de 29 de julio, de depósito legal. *Boletín Oficial del Estado*, 107, 5 de mayo de 2022. <https://www.boe.es/eli/es/l/2022/05/04/8> [Consulta: 14/06/2024]

Llueca Fonollosa, C. (2005). "Webs siempre accesibles: las bibliotecas nacionales y los depósitos digitales nacionales". *BiD: textos universitarios de biblioteconomía i documentació*, núm. 15 (diciembre). <https://bid.ub.edu/15lluec2.htm> [Consulta: 14/06/2024]

Parent, I., et al. (2021). Directrices UNESCO/PERSIST sobre selección del patrimonio digital para su conservación a largo plazo-2nd edición. IFLA repository. <https://repository.ifla.org/handle/123456789/1390> [Consulta: 14/06/2024]

Pastrana, A., y Cerdán, J. C. (2021). The Spanish Web Archive as a training field for Natural Language Processing models. IIPC netpreserve.org blog. <https://netpreserveblog.wordpress.com/2021/09/29/the-spanish-web-archive-as-a-training-field-for-natural-language-processing-models/> [Consulta: 14/06/2024]

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, de 25 de julio de 2015. <https://www.boe.es/eli/es/l/2011/07/29/23/con> [Consulta: 14/06/2024]

Van der Knijff, J. (2020). Restoring Liesbet's Virtual Home, a digital treasure from the early Dutch web (2023, 12 de diciembre). Bitsgalore: digital preservation – file format. <https://www.bitsgalore.org/2020/06/30/restoring-liesbets-virtual-home>

Webb, C. (2003). Directrices para la preservación del patrimonio digital. UNESCO, Information Society Division. https://unesdoc.unesco.org/ark:/48223/pf0000130071_spa [Consulta: 14/06/2024]

