
La conservación de todos los contenidos digitales no es necesaria: guardemos solo lo imprescindible de los medios

The preservation of all digital content is not necessary: save only the essential of the media

Juan Carlos MARCOS RECIO (1), María OLIVERA ZALDUA (2), Juan Miguel SÁNCHEZ VIGIL (2)

(1) Facultad de Ciencias de la Información de la UCM, Av. Complutense s/n 28040, jmarcos@ucm.es. (2) Facultad de Ciencias de la Documentación de la UCM, C/Santísima Trinidad 37 28010, molivera@ucm.es, jmvigil@ucm.es

Resumen

El ingente volumen de producción desde la llegada de la tecnología digital ha causado disparidad de criterios entre los documentalistas, archiveros y bibliotecarios a la hora de decidir qué contenidos digitales se deben conservar. ¿Se requerirá en el futuro alguno de los contenidos de Facebook? ¿Las imágenes de Pinterest, Instagram, deberán guardarse al completo? ¿Qué sucederá con los vídeos de YouTube, Vevo o Vimeo? ¿Y con los tuits del pájaro azul? ¿O los datos personales y de las empresas de LinkedIn? Este trabajo pretende una reflexión sobre los contenidos digitales, sobre la cantidad de documentos que se generan en las plataformas y diarios digitales, y sobre su conservación, a partir de propuestas ya en marcha como las de la Biblioteca Nacional de España o la Biblioteca del Congreso. Se requiere que las instituciones tomen parte en los contenidos que se generan cada día en las redes sociales, interpreten su valor y gestionen para el futuro dichos contenidos.

Palabras clave: Contenidos digitales. Conservación digital. Diarios digitales. Plataformas. Redes sociales. Biblioteca Nacional de España. Biblioteca del Congreso.

1. Introducción

A diferencia de lo que sucedía en los comienzos de la imprenta, donde el valor de los documentos se estimaba por el número más que por los contenidos, en el siglo XXI la producción es tan elevada que el ser humano se ve incapacitado para asumir todo lo que se crea, tenga o no valor, para la sociedad. La archivística, biblioteconomía, museología, documentación y otras ciencias han sido guardianes del saber durante siglos, pero se están planteando una nueva forma de analizar y conservar los contenidos digitales.

El almacenamiento ha crecido de tal forma que los espacios digitales pueden albergar una producción tan alta que en cada avance científico se vislumbra un horizonte halagüeño, pensando que todo el saber y no saber de la humanidad puede

Abstract

The enormous volume of production since the advent of digital technology has caused a disparity of criteria among documentalists, archivists and librarians when deciding which digital content should be preserved. Will any of the contents of Facebook be required in the future? Should Pinterest images, Instagram, be saved in full? What will happen to videos on YouTube, Vevo or Vimeo? And with the blue bird tweets? Or the personal and corporate data of LinkedIn? This work aims to reflect on digital content, on the amount of documents generated in digital platforms and newspapers, and on their conservation, based on proposals already underway such as those of the National Library of Spain or the Library of Congress. It is required that institutions take part in the content that is generated every day in social networks, interpret their value and manage such contents for the future.

Keywords: Digital content. Digital preservation. Digital newspapers. Platforms. Social networks. National Library of Spain. Library of Congress.

entrar en formatos tan diminutos que se evitará pensar en que lo fundamental no es tener todo, sino establecer pautas para llegar a lo verdaderamente necesario. La escasez de información agudizó el ingenio de los investigadores, docentes, periodistas, científicos, etc., que sacaban de las pocas fuentes el material necesario para dar crédito a sus teorías. Ahora, las dificultades son máximas por la dispersidad de los contenidos, la producción incontrolable y las escasas medidas de seguridad de lo que se produce en digital.

La parte positiva es que la tendencia cada vez mayor a tener los contenidos en abierto, creando un número reducido de directorios, gestionados por bibliotecas y centros de documentación, servirán para desechar aquellos aspectos superfluos que nada aportan a la civilización. La cuestión es: ¿Quién es el encargado de seleccionar

aquello que tiene valor? ¿Quién será el responsable de conservarlo? ¿De todos los contenidos digitales, algunos tienen más valor que otros? ¿Dónde está el inicio y el final del recorrido digital? ¿Las empresas públicas seguirán como en tiempos analógicos conservando en sus espacios físicos? ¿Y las compañías privadas guardarán los millones de fotografías que llegan a Instagram o los miles de horas de vídeo que cada día se suben a YouTube?

Este texto pretende aportar algunas reflexiones que puedan orientar a bibliotecarios y documentalistas en este campo. No hay soluciones definitivas porque estamos en una fase de experimentación que nos llevará a guardar lo que es imprescindible, si queremos que la información pertinente llegue a nuestros usuarios. Esta ha sido, por años, la respuesta que bibliotecas y centros de documentación han dado. Ahora que se ven desbordados a la hora de guardar contenidos digitales, están pensando en cambiar la estrategia del todo vale por el de aquello que aporta un valor presente y futuro a la sociedad, en todos los ámbitos: cultura, política, economía sociedad, etc. Desde aquí se propone la “Teoría del Equilibrio”, que consiste en trabajar con documentos y áreas que a lo largo de la civilización han resultado de extremo interés en los procesos creativos y de investigación. La propia selección de los contenidos ya es un proceso de equilibrio entre lo que genera recursos y lo que carece de valor. El equilibrio está en la manera en que se hacen las búsquedas y la toma de decisiones por parte de los documentalistas y bibliotecarios cuando gestionan el fondo y los contenidos que formarán parte del mismo. Frente a ella, la “Teoría de la Dispersidad” ofrece la posibilidad de contar con muchos documentos pero sin estar seleccionados ni catalogados por la imposibilidad material de hacerlo. Pasados los años de la especialización, que vinieron a demostrar el error más absoluto cuando se sabía mucho de una sola cosa y nada del resto, una propuesta de equilibrio es la que deben aportar los documentalistas tras un adecuado proceso de selección. Sólo así los usuarios entenderán al recibir los documentos, el valor que tienen, puesto que se han filtrado y se han obtenido los más pertinentes. Este proceso no es nuevo en la documentación, pero se requieren algunos matices desde la llegada de Internet y las redes sociales.

2. ¿Hay límites en la producción de contenidos en el entorno digital?

A lo largo de la civilización se ha demostrado que en contadas ocasiones se ha puesto límite a la producción. Motivos políticos, religiosos, académicos, etc., fueron en algún momento una

excepción, pero el ser humano ha gozado de libertad para crear contenidos y hacerlos llegar a la sociedad. A mediados del siglo XX es cuando empiezan a tener una preocupación los científicos, al darse cuenta de que la producción de artículos, libros e informes que existían eran tantos que no los podían asumir. Solicitan que los documentalistas gestionen esa producción y les apoyen a través de resúmenes documentales e informes *ad hoc* para poder avanzar en la ciencia. Esta situación se mantuvo, más o menos estable, hasta la llegada de Internet, en la que el crecimiento ha sido tan exagerado que ya no existe forma de administrar lo que se produce. Es cierto que la Inteligencia Artificial (IA) puede actuar de manera ágil, haciendo que determinadas materias tengan sus propios contenidos de manera precisa, pero en general, la propuesta de contenidos que se generan en Internet y las redes sociales es inasumible.

Antes de precisar los límites, conviene señalar qué tipos/prototipos de documentos se deben contemplar en los centros de documentación, especialmente, los relacionados con la comunicación, que es el área en el que está inscrita esta propuesta. La Escuela de Leyes de Cornell (1992) en su Regla 34 señala qué se puede considerar información valiosa para ser conservada:

Cualquier documento designado o información almacenada electrónicamente, incluidos escritos, dibujos, gráficos, cuadros, fotografías, grabaciones de sonido, imágenes y otros datos o compilaciones de datos, almacenados en cualquier medio desde el cual se pueda obtener información directamente o, si es necesario, después de la traducción por la parte que responde en una forma razonablemente utilizable.

Está claro que no solamente lo relacionado con temas de comunicación ofrece valor más allá de la actualidad con la que conviven los medios de comunicación cada día. Hay otras propuestas, también valiosas y necesarias, para el funcionamiento de la futura sociedad. Por ejemplo, todas las leyes que se han venido generando a lo largo del tiempo y que requerirán de una consulta en el futuro. También, los datos personales que cada individuo atesora o no, en su vida. Este aspecto es más difícil de resolver. Parecería que la IA lo solventa guardando todo lo que producimos en una memoria de nuestro equipo o en la nube. Pero no siempre el ser humano actúa con diligencia. Por precisar esta idea, señalar que algunos de los móviles de los ciudadanos atesoran cientos, miles de fotografías que van quedando en su memoria y de las que no se hacen copias de seguridad. Eso supone en ocasiones un riesgo que termina en la pérdida de muchas, o todas, si no se salvan a tiempo. Además, están los avances

significativos de los procesos técnicos de los equipos. Esto significa que cada vez que se cambian contenidos de uno a otro, no siempre se copian al cien por cien, con la consiguiente pérdida de información.

Puesto que la tecnología facilita algunos procesos, se puede señalar que la producción de Internet que se muestra en las siguientes líneas, será vieja y no tendrá casi valor cuando el sector termine de leerla. Porque es tan dinámica que lo que se gestiona ya no se mide en meses, ni días, ahora se contabiliza en capacidades de memoria, multiplicadas por segundo. Así, un reciente estudio de la revista *Forbes* (Marr, 2018) recogía datos difíciles de asumir por la inteligencia humana:

La cantidad de datos que producimos cada día es realmente alucinante. Hay 2.5 quintillones de bytes de datos creados cada día a nuestro ritmo actual, pero ese ritmo solo se está acelerando con el crecimiento de la Internet de las Cosas (IoT). Solo en los últimos dos años se generó el 90 por ciento de los datos en el mundo.

¡Vale la pena releerlo! Si bien es casi imposible envolver su mente en torno a estos números, reuní algunas de mis estadísticas favoritas para ayudar a ilustrar algunas de las formas en que creamos estas cantidades colosales de datos cada día. Es posible que las generaciones nacidas en lo digital sean capaces de asumir una parte de esa producción, pero se quedarán muchos datos por el camino, quizás no haya que preocuparse porque esos sean los de menor valor, pero ¿y si fueran los más importantes?

De acuerdo con la información publicada en *Forbes*, los datos son difíciles de asumir, pero son la realidad a la que se enfrenta la sociedad. La necesidad de saber, de conocimiento, incluso de diversión que se ha generado con las redes sociales hace que cada segundo se efectúen en Google unas 40.000 búsquedas, es decir, 3,5 billones de búsquedas por día. Esto indica que se vuelven a generar nuevos documentos que se subirán a Internet con otro contenido, el mismo, o parecido. De ahí, el valor del documentalista para separar y saber cuál de esas informaciones y/o documentos merecen pasar a otra etapa. Y no hay que olvidar que Google es solo un lugar de encuentro que asume el 77% de búsquedas, pero hay otros motores de búsqueda que también facilitan datos, y entre todos, señala Carr, suman cinco mil millones de búsquedas al día. Es posible que muchas de esas preguntas ¿Qué tiempo hace en Toronto? ¿Cuál es el río más caudaloso del mundo? ¿Qué lleva una buena sopa de pescado? O bien otras similares, no tengan un valor para el futuro, puesto que los resultados dependerán siempre del momento.

La realidad en cuanto a la producción de contenidos es tan cambiante que ya hay instituciones que ofrecen al segundo los cambios que se generan en cada red social, de forma que se sabe cuántos vídeos está recibiendo YouTube o Vimeo, cuántas fotos entran en Instagram, o cuántos “me gustas” recibe Facebook y así con el resto de redes sociales. Este sitio es uno de los que muestran el movimiento sin cesar de las redes sociales: www.dealsunny.com/mobile-usage-statistics/ (Figura 1).

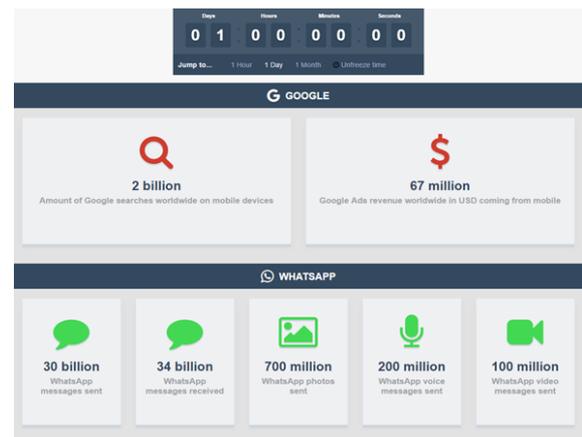


Figura 1. Cambios en Google en tiempo real

También conviene tener en cuenta algunos lugares que resumen de forma anual los datos globales generados a través de las redes sociales, como es el caso de *Data Never Sleeps* (en domo.com (2018)).

Esta situación se ha venido observando desde hace tiempo. Precisamente, esa fue la preocupación a finales del 2000 para dos investigadores Peter Lyman y Hal R. Varian que durante varios años estuvieron publicando el estudio: *How much information?* en el que evaluaban la producción mundial de contenidos, teniendo en cuenta que en aquel entonces las redes sociales estaban iniciándose.

En definitiva, no hay límites para la producción de contenidos, ni para contarlos, la cuestión a resolver es su gestión y su posible uso en el futuro, de manera que la información fluya ágil y eficaz. Para ello, habrá que pensar en guardar lo que valga para llegar antes a los contenidos decisivos. Mientras que las redes sociales crecen, los contenidos en los medios de comunicación digitales no han aumentado en la misma proporción de ahí que haya una brecha cada vez mayor. Los medios tienen capacidad tecnológica, pero la crisis ha reducido considerablemente las plantillas y con ellas la creación de contenido.

3. Plataformas y medios ante la desaparición de la información

La materia prima de los medios de comunicación es la información, que se convierte en documento cuando pasa a ser un referente para una sociedad. El proceso documental de los nuevos periódicos ha cambiado considerablemente en una década. Mientras que en el siglo XX se tomaban como referencias algunas publicaciones del centro de documentación para elaborar las noticias; en la actualidad se usan las informaciones como enlaces para completar la información, en algunos casos con otras informaciones ya publicadas por el redactor o de otros periodistas del mismo medio.

Sigue habiendo un proceso de análisis, de gestión y clasificación, pero el valor de la documentación está en el apoyo directo que hace el propio redactor sobre otras informaciones ya publicadas. Aquí, el reto está en gestionar los fondos que se van acumulando en digital, para qué sirven y por cuánto tiempo tendrán actualidad.

Existen ya proyectos que evalúan estas situaciones y piden que la administración del Estado se implique, pues hay ciudadanos que consideran que los medios de comunicación deben dar marcha atrás y despublicar informaciones cuando hay sentencias a su favor. En los medios tradicionales había un turno de réplica. Ahora se exige que se retiren los contenidos ya publicados y que se saquen de los centros algunas informaciones ya publicadas. Este sería el caso Gawker, un blog en el que tras declararse en bancarrota, salieron a subasta sus activos entre ellos sus nombres de dominio y casi 200.000 artículos archivados. Hubo otra parte que fueron comprados por una compañía de Univisión por 135 millones de dólares. El presunto comprador, Peter Thiel pretende borrar parte de su historia: “Thiel no ha dicho por qué quiere Gawker, aunque la posible adquisición le permitiría escribir historias sobre su vida personal que todavía están disponibles en el sitio web, y eliminar el alcance de nuevos litigios entre él y Gawker” (DiNapoli, 2018). Lo que pretendía el comprador era borrar después su historial en dicha publicación.

No es la primera plataforma y/o medio que ve peligrar sus fondos documentales. De hecho, se tuvo que crear el Proyecto de Periodismo Offshore <http://www.offshorejournalism.com/#/>, que consiste en enviar una señal de socorro a los medios que pueden perder para siempre su contenido. Este es un tema serio para los centros de documentación, pues se trata de la memoria colectiva que han generado por años los medios. Como se señala en su sitio web (Offshore Journalism, 2017):

El derecho a ser olvidado, las órdenes judiciales, la presión de los abogados y la censura absoluta por parte de la policía están haciendo agujeros en nuestra memoria colectiva. Decenas de artículos desaparecen cada mes y se pierden para siempre. Nuestra comprensión del pasado se verá dramáticamente obstaculizada por los fragmentos de contenido que desaparecieron de la web. ¿Qué sucede si un elemento clave de la siguiente biografía de su líder se eliminó del archivo de los periódicos de su país?

Resolver cuestiones como ésta llevaría a plantear situaciones complicadas como se propone en este Proyecto de crear vehículos en alta mar para guardar una copia del archivo del periódico. Es pronto para evaluar los resultados, pero los tiempos se acortan ante situaciones comprometidas para los contenidos que guardan los centros de los medios.

Un tercer ejemplo puede ayudar a entender este proceso. Se trata de una pequeña publicación italiana: PrimaDaNoi.it, que se vio obligada a retirar un artículo, eliminarlo de su centro de documentación y pagar 5.000 euros, como señala Owen (2018): “La Corte Suprema, que dictaminó en 2016 que las noticias tienen una fecha de vencimiento efectiva y que esta historia la había alcanzado”. Acontece que una vez pasado el tiempo de valor de la información, las personas incluidas en ella requieren mejorar su reputación, por eso se habla de tiempo ilimitado, incluso cuando el interés de los lectores ya no es grande. Este revés para algunas publicaciones rompe de manera radical con la manera en que se conserva la información, ya que los documentalistas no pueden estar revisando hacia atrás lo que ya se publicó. Parte de la solución pasa por guardar los archivos digitales de los medios en países donde la libertad de expresión sea más fuerte.

Existe una solución intermedia que no acaba de convencer a los editores. Se trata de lo que el citado Owen (2018) llama escribir para el futuro:

Significa proporcionar a cada elemento de información todos los atributos que lo harán posible y relevante para un público diferente, en un momento diferente. Ya sea que los editores y los periodistas sean o no plenamente conscientes de la posibilidad, este es un paso gigante en la redefinición de la libertad de expresión y la libertad de prensa.

Lo que resulta complicado es cómo se van a identificar los elementos informativos que el documentalista debe conservar para elaborar las futuras informaciones.

En la actualidad, la crisis de los medios por la escasa inversión publicitaria ha obligado a los editores a prescindir de parte del equipo de redacción. También a dar un paso más hasta hacer que el medio desaparezca. Así, en 2017, *Diario Digital* cerró según su editor por cuestiones

financieras. Fue el primer periódico en línea y llegó a su fin después de 17 años. En redacción había 80 personas. La conclusión inminente es: ¿Dónde han quedado esas informaciones? ¿Las guarda Archivo.pt? ¿Están en algún otro lugar? Al menos una parte del texto y de sus imágenes están conservadas en Archivo.pt.

Mientras, algunos medios desaparecen y con ellos los contenidos, a excepción de algunas instituciones que los conservan, por el momento. También hay que resolver el tema del derecho al olvido que plantean los ciudadanos y que obliga a los centros de documentación a borrar informaciones, puesto que se publicaron en un determinado momento y la situación del ciudadano ha cambiado respecto a los contenidos del archivo.

Una respuesta a todas estas iniciativas es Internet Archive <https://archive.org/>, una biblioteca sin fines de lucro que cuenta con millones de libros, películas, software, música, sitios web, etc. Incluye archivos sonoros, textos y documentos de televisión que garantizan la continuidad de los mismos, pero no están todos, simplemente una parte que se rastrea a través de la web. En el caso de la TV cuenta con 1.667.181 de las principales cadenas de Estados Unidos, pero también de otros países.

4. Bibliotecas y medios ponen límites a la conservación

Mientras que la mayoría de los medios de comunicación atraviesan situaciones complicadas a la hora de conservar y generar su patrimonio, por los nuevos cambios que se han producido con la llegada de la digitalización, las bibliotecas enfrentan retos diferentes a la hora de qué hacer con su patrimonio.

La primera consecuencia es resolver qué se debe o qué no debe conservar en los formatos digitales y en lo que se produce en las redes sociales. En consecuencia, ¿todo se debe conservar o sólo lo imprescindible? Y, ¿quién determina lo que es/no es imprescindible?

En este entorno del crecimiento exponencial de las publicaciones y contenidos digitales algunas bibliotecas y medios han llegado a acuerdos para poner límites. Por ejemplo, la *Library of Congress* no quiere ya los contenidos que se publican en Twitter. Según *La Vanguardia*, en 2010 la Biblioteca del Congreso de Estados Unidos firmó un acuerdo con Twitter para que le proporcionara cada día y cada uno de los tuits que se escribirían en el mundo. Además, retrospectivamente, desde el año 2006 que empezó a funcionar la compañía del pájaro azul, le envió todo el lote de lo que había publicado. Esta situación obligó a la biblioteca (Rodríguez, 2017):

A cambiar su política de almacenamiento [...] en primer lugar, el número de publicaciones se ha ido incrementando con el tiempo, a medida que se sumaban más usuarios y la actividad crecía: los 50 millones de tuits diarios que entonces producían 40 millones de usuarios se han convertido en más de 500 millones, con una base actual de 330 millones de usuarios. Asimismo la institución indica que los tuits le resultan cada vez menos útiles, ya que solo preserva texto.

Además, las capacidades de almacenamiento crecieron cuando la red social permitió que además de texto los usuarios pudieran publicar sus vídeos, fotos y gifs animados. Guardar los 12 primeros años completos de Twitter planteaba problemas, porque había autores que luego retiraban lo que habían publicado e iba en contra de una idea clara de la Biblioteca del Congreso, siempre selectiva en aquello que elige guardar para la posteridad. Este criterio selectivo es el que se defiende en este texto, ante tantos contenidos que apenas tienen valor y que ocultan los que de verdad deben estar listos para ser consultados.

Mientras que en las bibliotecas se pone freno a la conservación de contenidos web, en algunos medios se ha tratado de conservar sus contenidos, guardándolos en otros lugares sin ánimo de lucro y que están actuando de manera inmediata sobre algunas informaciones digitales, como es el caso de *Wayback Machine Explore* <https://archive.org/web/>, un programa de Internet Archive's con más de 347 mil millones de páginas web guardadas en el tiempo. Son más de 20 años de historial web accesibles a través de *Wayback Machine*. Se trabaja con más de 450 bibliotecas y otros socios a través de Archive-It. A modo de resumen estos son sus datos:

- 330 mil millones de páginas web
- 20 millones de libros y textos
- 4,5 millones de grabaciones de audio (incluyendo 180.000 conciertos en vivo)
- 4 millones de vídeos (incluyendo 1.6 millones de programas de noticias de televisión)
- 3 millones de imágenes
- 200.000 programas de software

Por su parte, *Gothamist Archive Retrieval Tool* <https://kpetrovi.ch/e-clips/> permite recuperar clips de los cachés de Google AMP. Los cachés caducan, por lo que esto solo puede funcionar por un día como mucho y habrá que conservarlos en el disco externo del usuario. La cuestión es que esta herramienta, necesaria para el periodista, fue cerrada recientemente, pero antes se envió una copia de sus datos a Archive.org evitando así

su desaparición. Tras despedir a 115 periodistas, el sitio web se quedó inactivo (Kissane, 2017):

Una fuente cercana a Gothamist / DNAInfo informa que los datos de los sitios se han conservado y que el editor de los sitios ha estado en contacto con el Internet Archive para mantener un archivo a largo plazo de los años de periodismo actualmente inaccesibles en Gothamist y sitios relacionados. Cuando se le preguntó sobre la eliminación del contenido de varios sitios, nuestra fuente dijo: 'No lo pensaron. La directiva era enviar cada página a esta declaración'.

En ocasiones, un cierre del medio posibilita la apertura de otro y su continuidad, como fue el caso de DANinfo <https://blockclubchicago.org/> que guardó su fondo hasta 2017, año de su desaparición. Como señala Bilton (2017), el proceso final es lograr que se archiven estos sitios web con contenidos periodísticos. Algunos son de corta vigencia, pero contienen ya datos interesantes para entender una parte de la sociedad norteamericana:

Archivo, archivo, archivo. Si bien, *The Times* ha confirmado que en DNAInfo se están realizando esfuerzos para archivar los sitios, también hay varios en cursos externos [...] Paul Ford creó una hoja de cálculo que enlaza con más de 50.000 artículos eliminados, aunque solo funciona con Gothamist.com en este momento.

El proceso final camina hacia una selección de contenidos, pero no está claro aún con qué herramientas se aplicarán.

Por lo que respecta a la Biblioteca Nacional de España y al igual que otras instituciones inició un proyecto en colaboración con Internet Archive en 2009 para archivar sitios web que incluyen blogs, foros, imágenes, vídeos, contenidos en Twitter, etc. En realidad son páginas web con el dominio .es y los subdominios .com, .edu, .gob, .org, etc., todo ello con el fin de conservar el patrimonio español de lo que se publica en Internet. En 2014 se comenzaron a hacer recolecciones con medios propios y ya en el 2015 se aprobó el Real Decreto que regula el Depósito Legal de las publicaciones en línea y la manera de hacer este depósito (Peguero; Ordás. 2019). Con posterioridad se puso en marcha el proyecto: "Servicio de apoyo a la gestión de colecciones del depósito legal de las publicaciones en línea de la Biblioteca Nacional de España", en marzo de 2017.

El proyecto, aun siendo exhaustivo, reconoce que en las fechas actuales no se puede alcanzar toda la producción que se genera y que se necesitará mejor tecnología y recursos para hacerlo (BNE, 2016):

Debido al enorme tamaño de Internet y a los medios tecnológicos de los que disponemos actualmente, a día de hoy es imposible aspirar a la exhaustividad en el archivado web. Por eso, para intentar guardar

la mayor cantidad posible de información web, la Biblioteca Nacional de España ha optado por un modelo mixto que combina recolecciones masivas y selectivas, como hacen otras bibliotecas nacionales del mundo.

La BNE no se plantea conservar todo sino una selección tomando como referencia las Directrices para la preservación del patrimonio digital de la UNESCO (2003) y la Recomendación de la Comisión Europea de 24 de agosto de 2006 sobre la digitalización y la accesibilidad en línea del material cultural y la conservación digital. Es un proyecto que ya está dando resultados y un primer rastreo de los documentos .es se hizo en 2016 durante tres meses, pero además se avanzó en la colaboración y cooperación en algunos centros de Comunidades Autónomas con el fin de establecer un Depósito Legal para las publicaciones en línea de maneja colaborativa. En este sentido, la BNE ha descentralizado algunos servicios y son otros centros y bibliotecas quienes gestionan ya sus propias colecciones web con herramientas construidas por la BNE.

Un aspecto importante es que al tener un elevado índice de producción con respecto, solo a los dominios .es, la propia BNE se planteó ser selectivos en algunos de sus contenidos. Se trataba de pensar en fondos que podrían tener un significado especial para el futuro del patrimonio de la institución y que se habían originado en páginas de Internet (BNE, 2016):

Biblioteca ha realizado desde entonces varias recolecciones selectivas sobre acontecimientos relevantes para la historia y la cultura españolas, como la muerte de Adolfo Suárez, la abdicación de Juan Carlos I, la proclamación de Felipe VI, las elecciones europeas de 2014, las locales y autonómicas de 2015 y las Elecciones Generales de 2015-2016.

Así, acontecimientos especiales requerirán un seguimiento mayor con respecto a todo lo que se publica en las redes y que aconseja sea preservado para el futuro, puesto que tiene un valor documental y patrimonial importante y será material de consulta casi seguro.

En esta tarea, la BNE no está sola. Cuenta con la comunidad científica e investigadora quien la facilita pautas para saber qué contenidos pueden tener una valor documental para el futuro. Para ello, durante 2018 y el año en curso se esperan digitalizar unas 133.000 páginas de unas 545 obras con el fin de dar valor y conservar aquellas publicaciones científicas significativas.

Por último, otro de los proyectos de recuperación y conservación activa del patrimonio de la BNE se ha puesto en marcha recientemente. Se trata de un proyecto colaborativo para mejorar y enriquecer las colecciones digitales y reutilizar los

recursos digitales ya existentes y catalogados. En este proyecto la BNE va de la mano de Red.es. En este caso, la reutilización es necesaria para que una colectividad aporte los datos necesarios para identificar lo que ya se conserva en digital: “ComunidadBNE trata de poner en contacto a la BNE y colaboradores virtuales para transcribir, geolocalizar, identificar y etiquetar personas en las imágenes digitales de sus fondos históricos o añadir información a los registros de su catálogo bibliográfico y de autoridades. A partir de hoy (06/02/2019) la plataforma está disponible para que cualquiera pueda contribuir, haciendo que el patrimonio de la Biblioteca sea más accesible, más rico, visible y recuperable. Y, con ello, a impulsar su utilización en la investigación, el desarrollo tecnológico o la educación (BNE, 2019). Se abre así la BNE a la comunidad que desee ser partícipe en la organización y gestión de fondos documentales digitalizados y lo que se pretende es mejorar los datos para que los usuarios tengan no solo acceso al documento, sino también las referencias que dan más valor al mismo.

5. Conclusiones y reflexiones

La aparición de redes sociales trajo un nuevo modelo de conservación. Hasta fecha reciente, la mayoría de las bibliotecas y centros de documentación conservaban todo lo que llegaba a su fondo. En ocasiones, alguna donación o compra obligaba a hacer un expurgo para presentar un fondo más completo. Cuando el crecimiento exponencial de producción de contenidos se hizo tan alto, algunas instituciones decidieron poner límites a todo lo que se guardaba.

El ser humano cree lo digital está en una nube y no ocupa lugar. Que lo que antes se guardaba en formatos digitales se puede subir a la nube y allí apenas es una muy pequeña parte de lo que se puede almacenar. También es cierto que las capacidades de almacenamiento se han multiplicado de tal forma que nada tiene que ver con lo que sucedía hace una década. Pero, no está resuelto el almacenamiento porque si bien crece de forma exponencial, los contenidos vienen duplicándose cada dos años. Y esto es imparable.

En este sentido, conviene reflexionar si todo lo que se produce tienen un valor actual y lo tendrá en un futuro, siendo conscientes de que no se sabe qué o qué no será lo que el ser humano necesite. Pero un tema está claro: evitar los contenidos, duplicados, triplicados, etc., porque no ayudarán en la gestión se los mismos en un futuro. La necesidad de información vendrá definida por la rapidez en su entrega, por la gestión eficaz y por la claridad con la que se haga al usuario.

Cuando se menciona que guardar lo imprescindible es la manera correcta, no se señala que haya que desechar contenidos de valor; lo que se está indicando es que como los contenidos serán tan iguales, habrá que determinar cuáles o cuáles no tendrán el privilegio de pasar a la historia. Al menos, en las bibliotecas y/o centros de documentación, porque hasta la actualidad, las redes sociales parece que aguantan y conservan todo lo que se produce. Es un reto pendiente que tienen las grandes empresas que ponen al usuario en el centro de su gestión, ya que su éxito está en proporción a los contenidos que suben. Cuantos más mejor, cuantos más seguidores, mejor, etc. Pero, bibliotecas y centros de documentación pueden afrontar el reto de no guardar más que aquello que sea imprescindible. Tener millones de documentos porque son fáciles de almacenar, no engrandece a un lugar; al contrario, porque cada vez más se demanda la información más precisa, pertinente y generada en el menor tiempo posible. En breve, las máquinas harán estas tareas de selección de contenidos y entonces descubriremos si guardar todo fue un éxito o deberíamos haber apostado por lo imprescindible.

Referencias

- Biblioteca Nacional de España (2016). Archivo de la Web Española. www.bne.es/es/Colecciones/ArchivoWeb/Historia/index.html (2019-03-12).
- Biblioteca Nacional de España (2018). La BNE digitalizará 545 obras a propuesta de la comunidad académica e investigadora. 17 de julio 2018. www.bne.es/es/AreaPrensa/noticias2018/0712-Digitalizacion-propuestas-investigadores.html (2019-03-15).
- Biblioteca Nacional de España (2019). La Biblioteca Nacional de España pone en marcha una plataforma de colaboración pública en la red. 6 de febrero 2019. www.bne.es/es/AreaPrensa/noticias2019/0206-presentacion-comunidad-bne.html (2019-03-10).
- Bilton, Ricardo (2017). The risk of billionaire-funded media, the importance of archiving, and other takeaways from the demise of DNAInfo and Gothamist. 03 de noviembre 2017. www.niemanlab.org/2017/11/the-risk-of-billionaire-funded-media-the-importance-of-archiving-and-other-takeaways-from-the-demise-of-dnainfo-and-gothamist/ (2019-03-10).
- Cornell Law School (1992). Federal Rules of Civil Procedure › Title V. Disclosures and Discovery › Rule 34. Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes. www.law.cornell.edu/rules/frcp/rule_34 (2019-03-12).
- DiNapoli, Jessica (2018). Peter Thiel presenta una oferta por Gawker. Enfrenta desafíos. Reuters 11 de enero 2018. www.reuters.com/article/us-gawker-thiel/peter-thiel-submits-bid-for-gawker-faces-challenges-idUSKBN1F02V2 (2019-03-12).
- Domo.com (2018). Data never sleeps 6.0. <https://www.domo.com/learn/data-never-sleeps-6>
- Kissane, Erin (2017). How to Save DNAInfo/Gothamist Bylines. 02 de noviembre 2017. <https://source.opennews.org/articles/save-bylines/> (2019-03-15).

Lyman, Peter; Varian, Hal R. (2000). How much information?.
<http://groups.ischool.berkeley.edu/archive/how-much-info/> (2019-3-20).

Marr, Bernard (2018). ¿Cuántos datos creamos cada día? Las estadísticas alucinantes que todos deberían leer. Forbes 21 de mayo 2018 www.forbes.com/sites/bernard-marr/ 2018/05/21/ how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#1a51d07060ba (2019-03-15).

Owen, Laura (2018). The Offshore Journalism Project would let newsrooms send a “distress signal” when their content is at risk of being lost forever. 08 de enero 2018. www.niemanlab.org/2018/01/the-offshore-journalism-project-would-let-newsrooms-send-a-distress-signal-when-their-content-is-at-risk-of-being-lost-forever/ (2019-03-15).

Enviado: 2019-03-27. Segunda versión: 2019-06-04.
Aceptado: 2019-06-20.
