

El Consorcio Internacional de Preservación de Internet

Tratando de asegurar el acceso a largo plazo a la memoria de la Red

LUIS CRESPO ARCÁ

Las nuevas tecnologías abren puertas a nuevos problemas que exigen una resolución inmediata pues, ellas mismas, son productos de una sociedad de consumo vertiginosa y, como hijas de la misma, de vida y uso pasajeros. Sin embargo, es conocido por la comunidad científica de archivistas o bibliotecarios que esta vida efímera no puede confundirse con que sean objetos/documentos de insuficiente interés pues son fiel reflejo de un modo de pensar y de vivir. Esta carrera mediática e informativa obliga a la creación y perma-

técnicos en su traducción al español van por detrás de su evolución, de ahí, por ejemplo, el uso frecuente de acrónimos, tan del gusto del mundo anglosajón, que definen sistemas y herramientas de procesado y almacenamiento de información, etc.

El proceso de archivar la Red es simple en su formulación: se trata tan solo de capturar sitios y archivarlos. La preservación, en este contexto, implica conseguir la capacidad de hacer accesible la información para el futuro con un sentido completo de sus valores. Desde que se tomó conciencia



nente actualización de profesionales –individuales o en grupos de trabajo– especializados en este nuevo campo de la Archivística. Como tantas veces en lo referente a la tecnología, también los términos y conceptos

de esta necesidad –algo que no surgió hasta pasados unos años de su creación y uso generalizado– la evolución ha sido constante. Este proceso ha implicado tanto el esclarecimiento de qué se quiere guardar –un sitio web, una página o una parte de un sitio web para que sean de acceso a futuros historiadores, investigadores o público general– hasta los sistemas (*software*) que sean capaces de permitir el acceso a esos documentos creados con programas informáticos disímiles. Para llegar a este punto de preservación, el primer paso obligado es el de ser capaces de realizar una captura sistemática de aquellas páginas que se desean guardar. Sin embargo, dada la complejidad técnica de este requisito (incluido los sistemas de almacenamiento, uso compartido entre instituciones, programas de acceso libre...) se puede afirmar que los esfuerzos actuales están aún

en una fase de búsqueda y maduración.

Para estas instituciones son fuente básica de información los programas y empresas que gestionan los llamados “araña web” que se dedican a inspeccionar y almacenar las páginas de la Red de forma metódica y automatizada. Uno de los usos más frecuentes que se les da consiste en crear una copia de todas las páginas webs visitadas para su procesamiento posterior por un motor de búsqueda que indexa las páginas proporcionando un sistema de búsquedas rápido. Las arañas webs comienzan visitando una lista de URL, identifica los hiperenlaces en dichas pági-

nas y los añade a la lista de URL a visitar de manera recurrente según un determinado conjunto de reglas.

En el año 2003 se creó el Consorcio Internacional de Preservación de Internet (*International Internet Preservation Consortium*, o IIPC, en inglés) formado con instituciones que tienen objetivos similares a la hora de preservar el contenido de la Red con fines de patrimonio cultural y que comparten sistemas y herramientas similares de recolección y acceso. El grupo lo constituyen archivos, bibliotecas y proveedores de servicios informáticos. Su comité directivo creó en el 2007 un grupo de trabajo de preservación (*Preservation Working Group*, o PWG, en inglés) con los siguientes mandatos: identificar sitios de la Red a fin de hallar qué sistemas y prácticas ya se están empleando en otros sitios, viendo si tienen utilidad y cómo pueden ser usados para archivar sitios así como identificar sus lagunas técnicas; hacer recomendaciones para la mejora de las herramientas, estándares, prácticas, directrices, ensayos y las posibles nuevas líneas de investigación; finalmente, promover el reconocimiento de aquellos requisitos técnicos necesarios para preservar los recursos de la Red que no hayan sido conseguidos por otros programas de preservación y/o emprendedores activos.

Los retos a los que se enfrenta este grupo de trabajo implican desarrollar unos estándares y hallar las mejores prácticas posibles. Actualmente tan solo existe una pequeña cantidad de estándares específicos para archivar la Red, pero estos son críticos. La mayoría de instituciones usan, o han usado, el llamado formato ARC, formato de compresión de datos sin pérdida, que también da nombre al programa que los genera, pero que tiene la pega de no poder comprimir directorios enteros de las páginas y asegurar su acceso abierto (Nintendo, por

ejemplo, usa este formato para sus juegos del *GameCube* o la *Wii*). Este formato está siendo desplazado por el llamado WARC (*Web Archive*) –consiste en un método para combinar múltiples recursos digitales en un archivo de archivos, agregado junto con la información relacionada– que permite una mejor recolección y acceso a los datos y facilita el intercambio entre instituciones.

Además de promover y apoyar el desarrollo de herramientas informáticas, el grupo de preservación ha hecho trabajos de investigación en diversos aspectos tales como la migración o la emulación de contenidos digitales para saber si se ajustan a la necesidad de cómo archivar los contenidos de la Red y para identificar la distancia entre las soluciones y posibilidades existentes y las deseables por cualquier archivero o bibliotecario. El principal problema de la preservación digital reside en poder permitir el acceso a la información de la forma más auténtica posible. A día de hoy, no existe una estrategia que sea obvia para conseguir este fin de cara al futuro con los archivos digitales. Las habituales estrategias de la migración o la emulación pueden no ser adecuadas para los archivos de la Red en el mismo modo que sí lo son para otros archivos digitales. Cada institución que almacene archivos digitales debe lidiar con el desafío de encontrar la mejor solución para sus fondos/colecciones concretas. El grupo de preservación del Consorcio Internacional de Preservación de Internet mantiene reuniones regulares justamente para saber de las necesidades y técnicas de migración y emulación desarrolladas por cada institución en función de sus necesidades y de los diferentes factores que las condicionan. De todo ello esperan ir hallando soluciones a los problemas existentes y a los que irán surgiendo en el futuro inmediato. ■

