

El historiador y la historia en la Edad Oscura Digital

José Ramón Cruz Mundet

Universidad Carlos III de Madrid
jrcomar@bib.uc3m.es

Resumen: El mundo se hace digital y con ello los documentos, fuentes de información para la historia. Nuevos soportes y medios cuya conservación permanente no está asegurada, lo que ha llevado a predecir la venida de una Edad Oscura Digital, en la que se disiparían la mayoría de las fuentes para la historia. En ese contexto se analiza la presencia del historiador en el debate, las nuevas formas de hacer historia —que podrían peligrar— y se pretende vislumbrar el futuro de los documentos digitales como materia prima para la historia.

Palabras clave: documentos electrónicos, preservación digital, fuentes de información, Tecnologías de la Información y de las Comunicaciones (TIC), archivística.

Abstract: The world becomes digital and its documents become information sources for history. The permanent conservation of new media outlets and supporting platforms is by no means assured. The possibility of their disappearance has led many to predict the coming of a Digital Dark Age in which the majority of sources used for the historical study would disappear. Given this context, we analyse the historian's presence in this debate and the new ways of practicing history that could be jeopardized. This article offers a glimpse into the future of digital documents as raw material for history.

Keywords: electronic records, digital preservation, information sources, information and communication technolgis (ict), archival science.

Introducción

En 1995 Jeff Rothenberg, renombrado experto en tecnologías de la información, iniciaba uno de sus artículos más populares¹ con la siguiente anécdota, un clásico ahora en preservación digital:

«Es el año 2045 y mis nietos están explorando el desván de mi casa, cuando encuentran una carta fechada en 1995 y un CD-ROM. En la carta se dice que el disco contiene el documento que proporciona la clave para acceder a mi fortuna. Mis nietos están comprensiblemente emocionados, pero nunca han visto antes un CD —excepto en las películas antiguas— e incluso si pueden encontrar una unidad de disco adecuada. ¿Cómo se ejecutará el *software* necesario para interpretar la información del disco? ¿Cómo pueden leer mi documento digital obsoleto?».

Así ilustraba una preocupación que permanece sin respuesta treinta años después. Una cuestión entonces ineludible y hoy perentoria, en la que el transcurso del tiempo es una amenaza.

Desde la aparición de la fotografía en los años treinta del siglo XIX se han sucedido nuevos medios y soportes documentales (como el microfilme, el cine, el disco, la cinta magnética...) que han convivido con el papel (manuscrito o impreso) como vehículo de la información. A partir de mediados del siglo pasado se inició un cambio de tendencia con la irrupción, lenta al principio, acelerada en las últimas décadas, de nuevos sistemas que han ido desplazando a los medios analógicos, cuya sustitución se presenta como horizonte ineludible; se trata de las Tecnologías de la Información y de las Comunicaciones (TIC).

El mundo se hace digital, la economía se digitaliza, como lo hacen las organizaciones y los individuos. Salvo el libro y la lectura que persisten en el papel —al menos en parte—, las administraciones, las empresas y los ciudadanos avanzamos hacia el cercano horizonte del gobierno electrónico, de la sociedad digital. Como muestra un botón: el 2 de octubre de 2016 entraban en vigor las Leyes 39/2015, del Procedimiento Administrativo Común, y

¹ Jeff ROTHENBERG: «Ensuring the Longevity of Digital Information», *Scientific American*, 272, 1 (1995), pp. 42-47.

40/2015, de Régimen Jurídico del Sector Público, orientadas ambas a ordenar una Administración totalmente electrónica, cuyas actividades se basen exclusivamente en documentos digitales, en plazo inferior a un año. Un panorama armonizado en prácticamente todo el mundo, que avanza hacia una segunda globalización, la digital.

La Edad Oscura Digital

La cara de la digitalización se dibuja con un sinnúmero de ventajas: rapidez, eficiencia, economía, exhaustividad, precisión..., en cuantos sectores y actividades se aplique; la cruz, por lo que aquí nos interesa, es, más que una sospecha, una certeza: la desaparición en buena medida de la información, de los documentos de nuestro tiempo, algo que ya viene sucediendo y que se está dando en el corto plazo.

La preocupación de los expertos y las primeras voces de alarma se dieron en los inicios de la revolución tecnológica. En 1997, cuando los usuarios de Internet² apenas representaban el 2 por 100 de la población mundial y las conexiones eran vía módem, Terry Kuny, consultor tecnológico canadiense, acuñó la expresión *Digital Dark Age* (Edad Oscura Digital) en su alocución dirigida a la conferencia de la Federación Internacional de Asociaciones Bibliotecarias³. En ella sostenía que estábamos —y continuamos— en una era en la que se valoran más el cambio y la velocidad que la conservación y la longevidad.

Una inquietud que no ha parado de crecer a medida que se ha constatado la pérdida de cantidades difíciles de precisar de docu-

² Frente al 44 por 100 del año 2015, según la Unión Internacional de Telecomunicaciones. Véase *Informe sobre el Desarrollo Mundial/TIC de las Telecomunicaciones y estimaciones del Banco Mundial*, disponible en <http://datos.bancomundial.org/indicador/IT.NET.USER.P2> (consultado el 25 de febrero de 2017).

³ Terry KUNY: «A Digital Dark Ages? Challenges in the Preservation of Electronic Information», en *63th. IFLA Council and General Conference. Workshop: Audiovisual and Multimedia joint with Preservation and Conservation, Information Technology, Library Buildings and Equipment, and the PAC Core Programme, September 4, 1997*, 12 pp., disponible en http://www.academia.edu/19711051/A_Digital_Dark_Ages_Challenges_in_the_Preservation_of_Electronic_Information (consultado el 16 de enero de 2017).

mentos y de fuentes de información. El último en sostener esta línea argumental, Vinton Cerf, uno de los padres de Internet y vicepresidente de Google, con motivo de la *Decentralized Web Summit*⁴ en junio de 2016, se declaraba preocupado porque el contenido digital será cada vez menos accesible y «si dentro de cien años la imagen digital de nuestra sociedad no es accesible, seremos un enigma en el siglo XXII»⁵.

Este desvanecimiento es consecuencia de la obsolescencia tecnológica, un factor determinante y convertido en lugar común, pero no el único. Se trata, en efecto, de medios de duración incierta dentro del corto plazo (así, los discos duros comerciales tienen solo garantía de un año y los profesionales de cinco⁶, pues tales son los límites a partir de los cuales comienzan a estropearse). Las generaciones tecnológicas se suceden en plazos cada vez más cortos, el *hardware* y los sistemas operativos desaparecen, y con ellos los programas ejecutados, junto con los documentos producidos. La seguridad es otro de los puntos débiles que se encuentra a la orden del día y toca en todas las facetas de la tecnología.

A estos viejos conocidos se suman otros, como el riesgo de fallos que provocan pérdidas de información⁷ y se sitúa entre el 1 y el 13 por 100, en función de la tipología, con una media del 3 por 100. ¿Quién no ha perdido imágenes y textos con el mensaje de *Fatal error. File not found* o similar? Asimismo, las TIC dependen de normas y estándares para su funcionamiento que se suceden y superponen sin fin⁸, propiciando la diversidad y con ello la incompatibilidad entre sistemas. La autenticidad de los documentos⁹ es

⁴ Disponible en <http://www.decentralizedweb.net>.

⁵ «Father of the Internet Worries Our Digital History Is Disappearing», disponible en <http://europe.newsweek.com/fatherinternetworriesourdigitalhistorydisappearing468642?rm=eu> (consultado el 16 de enero de 2017).

⁶ Véase el informe de la compañía Blackblaze, disponible en <http://blog.backblaze.com/2013/11/12/how-long-do-disk-drives-last/> (consultado el 16 de enero de 2017).

⁷ Richard WRIGHT, Ant MILLER y Matthew ADDIS: «The Significance of Storage in the Cost of Risk of Digital Preservation», *The International Journal of Digital Curation*, 3, 4 (2009), p. 106.

⁸ Raivo RUUSALEPP *et al.*: «Standards Alignment», en Nancy MCGOVERN y Katherine SKINNER (eds.): *Aligning National Approaches to Digital Preservation*, Atlanta, GA Educopia Institute, 2012, p. 117.

⁹ Seamus ROSS: «Digital Preservation, Archival Science and Methodological

otro de los factores determinantes, teniendo en cuenta que el concepto analógico de original no es aplicable ni se pueden mantener en el ser y estado con que se crearon, pues la conservación implica el cambio de soporte, de *software*, de formato... y esto afecta a la integridad, a la fiabilidad y a la accesibilidad de los documentos. Características que es posible preservar, pero con incertidumbre, por los riesgos inherentes a las TIC y porque no hay experiencias de largo plazo. Sin olvidar, entre los riesgos, el de la pervivencia de las empresas que dan servicio de alojamiento (nube), de lo que sobran casos de dominio público como Megaupload, Ubuntu one, Fotopedia... en los que los usuarios han visto perdidos sus archivos.

La idea misma de nube resulta engañosa, ya que señala a lo inmaterial, a lo etéreo. Nada más lejos de la realidad. Basta con disipar la apariencia para ver que detrás hay centros de proceso y almacenamiento de datos con cantidades enormes de servidores informáticos ocupando hectáreas de terreno. Telefónica¹⁰ inauguró en 2013 cerca de Madrid uno de los más grandes del mundo, cuya superficie equivalía a ocho campos de fútbol. Este tipo de centros tiene, además de su construcción y equipamiento, unos costes de mantenimiento en los que a la renovación y la sustitución de los elementos tecnológicos se une un consumo desorbitado de energía eléctrica para la climatización, ya que debe mantenerse toda la infraestructura informática a una temperatura entre los 21 y 23° C. La preservación digital es muy costosa, porque además se trata de conservar bits de forma comprensible y utilizable, garantizando su accesibilidad en el tiempo con independencia de los cambios en la tecnología¹¹, de los cientos de formatos¹² vigentes y desaparecidos, y de mantener el carácter dinámico de los documentos, con los hiperenlaces actualizados.

Además de costosa, la preservación digital es compleja por la gestión de los derechos de *copyright* de los medios y de la informa-

Foundations for Digital Libraries», *New Review of Information Networking*, 17 (2012), pp. 43-68.

¹⁰ Véase http://www.tendencias21.net/telefonica/Telefonica-inaugura-su-mayor-centro-de-datos-que-albergara-los-servicios-digitales-mas-avanzados_a596.html (consultado el 16 de enero de 2017).

¹¹ José Ramón CRUZ MUNDET y Carmen Díez CARRERA: *Los costes de la preservación digital permanente*, Gijón, Trea, 2015.

¹² Véase http://www.digitalpreservation.gov/formats/fdd/browse_list.shtml (consultado el 16 de enero de 2017).

ción. Incluso cuando esos productos han caducado, y más allá de la existencia de las empresas que los han producido, los derechos siguen vigentes. Téngase en cuenta que en el mundo del *software*, de las publicaciones electrónicas, etc., se adquiere el derecho de uso, pero no la propiedad. Aunque se trate de contenidos de libre acceso (*open access*), hace falta recabar el permiso del propietario. Y aquí está el quid de la cuestión: lo señaló Cerf, cómo gestionarlo y pagar por el contenido y el *software* en periodos largos de tiempo. Como consecuencia de la lucha contra la piratería y del régimen de oligopolio de las publicaciones científicas, el régimen de licencias y de propiedad intelectual es cada vez más restrictivo y ello hará difícil, si no imposible, preservar información relevante. Por otra parte, si no lo hacen los poderes públicos, las empresas no están interesadas necesariamente ni obligadas legalmente a conservar a largo plazo si no obtienen un beneficio de ello.

Por último, además de costosa y compleja, la conservación digital es inabarcable por el ritmo de crecimiento de la información susceptible de conservación, que se duplica cada dos años¹³. De modo que hemos pasado de los 0,3 Zetabytes (Zb) producidos en 2007 a 4,4 en 2013; en 2017 el universo digital representará 16 Zb y se calcula que en 2020 alcanzará los 44 Zb.

Fiat lux

En esa carrera contra el tiempo y por la conservación, instituciones y grupos de investigación vienen trabajando desde comienzos de los noventa en la búsqueda de soluciones a los diferentes retos planteados, alcanzando tal cantidad que ha llegado a acuñarse la expresión «jungla de la preservación digital»¹⁴. Una carrera en la que la mayoría de los esfuerzos ha fallecido por caducidad y, sobre todo, a causa de la crisis económica que venimos arrastrando desde 2007, y ha mermado la sostenibilidad de mu-

¹³ Según el informe de IDC (2014), disponible en <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm> (consultado el 16 de enero de 2017).

¹⁴ José Ramón CRUZ MUNDET y Carmen Díez CARRERA: *Los costes de la preservación...*, p. 73.

chos de los proyectos. No obstante, hay realidades sólidas cuya permanencia parece asegurada, como el National Digital Information Infrastructure and Preservation Program¹⁵ (NDIIP), en el que trabajan cerca de 250 organizaciones de 48 estados y 26 países; la Alliance for Permanent Access¹⁶ (APA), creada en 2005 por un grupo de instituciones europeas de investigación científica para compartir recursos y cooperar en la creación de una infraestructura estable para el acceso permanente a la información científica digital; la Digital Preservation Coalition¹⁷ (DPC), una organización no gubernamental centrada en compartir conocimiento y recursos en materia de accesibilidad y preservación a largo plazo de información digital; la Open Planets Foundation (OPF)¹⁸, creada en el año 2010 con financiación de la UE, agrupa a las principales bibliotecas nacionales y de investigación, archivos nacionales, universidades y empresas tecnológicas, para constituir una comunidad de prácticas para la preservación a largo plazo del contenido digital; la LOCKSS Alliance (Lots Of Copies Keep Stuff Safe), primera de una serie de redes de preservación cooperativa basadas en soluciones distribuidas, o el Internet Archive, una suerte de arca de Noé de cuanto contenido libre discurre o ha discurrido por la red y merece ser conservado para la historia: texto, audio, imagen en movimiento, *software* y páginas web. Trabaja para prevenir que Internet y otros materiales digitales desaparezcan, vencer su carácter efímero y construir un repositorio digital que ayude a proteger el derecho a saber y a recordar. A los veinte años de su inauguración contiene más de 284.000 millones de páginas web y más de 510.000 millones de objetos digitales, creciendo a un ritmo de más de 500 millones de páginas web semanales¹⁹.

En el terreno estricto de los archivos²⁰, las aportaciones más sólidas se dan en el ámbito nacional, con Estados Unidos a la cabeza, además de pionero, pues ya desde los años sesenta vio la necesidad

¹⁵ Véase <http://www.digitalpreservation.gov/>.

¹⁶ Véase <http://www.alliancepermanentaccess.org/>.

¹⁷ Véase <http://www.dpconline.org>.

¹⁸ Véase <http://www.openplanetsfoundation.org/>.

¹⁹ Véase <https://archive.org/>.

²⁰ José Ramón CRUZ MUNDET: «Estrategias de preservación digital permanente en los archivos nacionales: un estudio comparativo», *Boletín de ANABAD*, 65, 3 (2015), pp. 127-148.

de garantizar la conservación permanente de los nuevos documentos electrónicos, de modo que el programa ERA (Electronic Records Archive) de NARA (National Archives and Records Administration) constituye la iniciativa más longeva del mundo. También los programas de los National Archives del Reino Unido, como los de Australia, se cuentan entre los mejores ejemplos de compromiso con la configuración de un patrimonio documental digital.

La información digital, como producto de las TIC, requiere normas para su creación, uso y conservación. En tal sentido se han desarrollado un buen número de ellas, como el modelo de referencia OASIS (Sistema de Información de Archivo Abierto), METS y PREMIS (metadatos), la arquitectura de referencia TOGAF y muchas otras que tiene por objeto establecer un conjunto regulado de buenas prácticas (aunque la abundancia y superposición comprometa su manejo)²¹.

Estas soluciones han surgido mayoritariamente de instituciones públicas, responsables de la conservación del patrimonio histórico, y están sostenidas con fondos públicos; pocas son las iniciativas participadas, siquiera, por empresas con intereses comerciales en el sector. Parece que la preservación digital no se percibe, por el momento, como una responsabilidad ni como un negocio rentable.

Lo que aún resta por alcanzar es lo más difícil, la solución a la caducidad, a la denominada obsolescencia digital; lograr soportes y programas que permanezcan estables y así mantener accesible la información en el tiempo. En laboratorio se han ensayado y continúan probando diferentes soluciones. Desde hace décadas se viene trabajando con el cristal de cuarzo como soporte de alta densidad, estable y legible a través de microscopio electrónico. Últimamente se investiga el ADN, por su capacidad de transmisión demostrada en restos óseos de animales de hace cientos de miles de años, cuyo código genético se ha podido descifrar; pero el ADN es muy inestable en condiciones normales de humedad y temperatura, necesita ser encapsulado en un medio estable como el cristal (cuarzo, sílice...), y la grabación de datos es costosa, pues se debe pasar del código binario

²¹ José Ramón CRUZ MUNDET y Carmen DÍEZ CARRERA: *Los costes de la preservación...*, pp. 87 y ss.

al genético (que es cuaternario)²², de modo que estamos lejos de ese añorado soporte estable a un precio asumible.

Hay proyectos como OLIVE²³ para la conservación de contenido ejecutable (juegos, programas...), el cual depende de la disponibilidad de sistemas operativos, códigos fuente y arquitecturas, de conservarlos junto con información del contexto de uso y derechos de propiedad intelectual.

También *Software Heritage*²⁴, proyecto que trata de archivar todo el *software* de código abierto del mundo, cuenta con más de 3.000 millones de archivos pertenecientes a 49 millones de proyectos. El objetivo es crear una colección de todo el código disponible para que se siga pudiendo acceder a él a largo plazo. Que cualquiera pueda tener acceso al mayor repositorio de código fuente de proyectos *open source* del mundo no solo para investigar y analizar, sino para buscar vulnerabilidades o crear aplicaciones aún mejores basadas en las existentes.

Además, la nube se materializa en soportes de almacenamiento que han variado poco: discos duros, discos ópticos durables y de alta densidad (como Blue-Ray, M-DISC...) y cintas magnéticas, a los que cabe añadir ensayos como el disco Rosetta de alta densidad que permite su lectura con un microscopio de mil aumentos y tiene una durabilidad prevista entre dos y diez mil años²⁵. Lo que está sin resolver es cómo hacer comprensibles e interpretables esos códigos dentro de cientos de años.

En fin, existen soluciones teóricamente impecables, pero que no es posible llevar a efecto; otras que funcionan en laboratorio, pero no son aplicables en un entorno sostenible, mucho menos rentable y competitivo. Al menos por ahora, la conservación digital permanente tiene más incógnitas que certezas.

²² «How to Store your Data for a Million Years», *Fast Company*, disponible en <https://www.fastcompany.com/3045215/how-to-store-your-data-for-a-million-years> (consultado el 16 de enero de 2017).

²³ De la Carnegie Mellon University e IBM (<https://olivearchive.org/>).

²⁴ Véase <https://www.softwareheritage.org/>.

²⁵ Long Tien NGUYEN y Alan KAY: «The Cuneiform Tablets of 2015», en *Onward! Essays track at SPLASH 2015*, Pittsburgh, PA, 2015, p. 3.

El historiador y la historia digital

Podría afirmarse que el concepto de historia digital ocupa un lugar limitado, lateral, en el debate historiográfico, de modo que es un tanto complicado aventurarse en uno por la escasa producción y, sobre todo, por la falta de autoridades. Conscientes del riesgo asumido, podríamos entender la historia digital como el proceso de investigación del historiador a partir de fuentes de información digitales y digitalizadas, mediante el empleo de técnicas y herramientas tecnológicas (informáticas) para el tratamiento de la información (extracción, organización, análisis y representación). Los resultados se hacen accesibles en forma de productos tecnológicos (webs, aplicaciones...) y también pueden materializarse, al menos en parte, a través de texto.

Hasta ahora, la historia digital presenta tres momentos²⁶: el primero basado en el uso de documentos y publicaciones digitalizados de los archivos, bibliotecas, etc.; el segundo es interdisciplinar y se caracteriza por incorporar otros medios, como los sistemas de información geográfica, y un tercero basado en el uso de datos y documentos electrónicos, con herramientas de explotación y de visualización, y algoritmos.

Cuando nos planteamos realizar este ensayo buscábamos identificar la relación entre los historiadores y las fuentes digitales para la historia; un binomio en el que era difícil hallar referentes. Llegados a este punto, y a partir de una averiguación informal entre contemporaneístas españoles, pudimos apreciar su aparente ausencia como resultado, por un lado, de una escasa disponibilidad de fuentes de información electrónicas y, por otro, concretamente en España, la elaboración de su historia contemporánea demanda todavía, y es prioritario lograr, el acceso a los fondos en papel, y analógicos en general, de épocas pretéritas, vetados a la consulta y que se remontan a hace ochenta y más años. Así las cosas, era y es difícil concebir que fondos digitales interesantes para la investigación

²⁶ Douglas SEEFELDT y William G. THOMAS: «What Is Digital History?», en *Perspectives on History*, 2009, disponible en <https://www.historians.org/publications-and-directories/perspectives-on-history/may-2009/intersections-history-and-new-media/what-is-digital-history#> (consultado el 20 de enero de 2017).

vayan a estar accesibles antes de echarse a perder, de ahí que mal puedan ser objeto de preocupación.

Rastreando la producción bibliográfica global pudimos constatar la existencia de una querencia general por el estudio de las propias tecnologías y de sus múltiples aplicaciones a la investigación, pero no tanto de obras orientadas a su uso y, sobre todo, a su empleo de forma diferente a como vienen siéndolo los documentos analógicos. Entre los primeros en tocar el tema, el norteamericano Seamus Ross²⁷ señaló en 1998 que pocos estudios se abordaban desde la perspectiva del historiador. En ese mismo año tenía lugar en Toledo²⁸ un congreso sobre los nuevos espacios de la historia, sin que ninguna de las intervenciones hicieran al caso; poco después, Fernández Izquierdo²⁹ publicaba un ensayo acerca de la aplicación de las tecnologías por parte de la investigación histórica desde los años sesenta, con una recopilación de proyectos; siempre desde la perspectiva del uso de la tecnología, no de la explotación de los nuevos materiales. Habrá que esperar para observar cierto cambio de tendencia, si bien incipiente y con toques de atención acerca del escaso interés de los historiadores por la investigación digital³⁰. Es en el ámbito de la historia política contemporánea, entre los investigadores centrados en los años noventa en adelante, donde el uso de fuentes digitales resulta inevitable³¹.

Junto con la relativa atención al tema, otra perspectiva visible es la brecha entre archiveros e historiadores al respecto de

²⁷ Seamus ROSS: «The Expanding World of Electronic Information and the Past's Future», en Edward HIGGS (ed.): *History and Electronic Artefacts*, Oxford, Clarendon Press, 1998, pp. 5-28.

²⁸ Francisco José ARANDA PÉREZ, Porfirio SANZ CAMAÑES y Francisco FERNÁNDEZ IZQUIERDO (coords.): *La historia en una nueva frontera. History in a New Frontier*, Cuenca, Ediciones de la Universidad de Castilla-La Mancha, 2000, p. 426.

²⁹ Francisco FERNÁNDEZ IZQUIERDO: «La historia moderna y nuevas tecnologías de la información y las comunicaciones», *Cuadernos de Historia Moderna*, 24 (2000), pp. 11-31.

³⁰ Alexander MAXWELL: «Digital Archives and History Research: Feedback from an End-User», *Library Review*, 59, 1 (2010), pp. 24-39.

³¹ «Political History in the Digital Age: The challenges of archiving and analysing born digital sources», en *London School of Economics and Political Science, Impact Blog*, disponible en <http://blogs.lse.ac.uk/impactofsocialsciences/2016/03/31/political-history-in-the-digital-age-born-digital-sources/> (consultado el 20 de enero de 2017).

lo digital, donde es ilustrativo el diálogo entre Blouin y Rosenberg³² acerca de las relaciones entre ambas disciplinas. La obra recoge, además de los retos de interés general que presenta la conservación, otros como: el iter documental y la desaparición de los conceptos de borrador, notas, redacciones...; la autenticidad y el contexto de los documentos³³, y, sobre todo, la selección, salida ineludible a la hiperinflación documental; práctica archivística tan controvertida para el historiador como eficaz para asegurar un patrimonio digital sostenible.

Una inercia, en definitiva, difícil de vencer y en la que predomina, entre quienes se han adentrado en la materia, la descripción de la tecnología como tendencia de historiadores y humanistas digitales más interesados en el medio (TIC) que en el objetivo (hacer historia). Llama la atención que desde la irrupción de las TIC los investigadores que se dedican a ello se orienten a la descripción de los nuevos medios, a glosar sus ventajas, a elaborar demostraciones sobre su potencial, pero no tanto a explotarlos y obtener resultados en forma de historia³⁴.

Asimismo hemos podido constatar una tendencia global al uso de documentos y publicaciones digitalizados por igual e indistintamente en todas las especialidades (desde la Antigüedad a la Contemporánea), pero se trata de eso, de materiales analógicos digitalizados para facilitar su acceso. Los ejemplos abundan tanto en los portales web de los archivos y bibliotecas como en forma de repositorios especializados³⁵, pero ni la naturaleza de la información original ni el abordaje de los investigadores varían respecto de sus originales. Y eso a pesar de que estas nuevas fuentes, tras adecuado proceso, permiten explotarlas de formas hasta hace poco impensa-

³² Francis X. BLOUIN jr. y William G. ROSENBERG: *Processing the Past. Contesting Authority in History and the Archives*, Oxford, Oxford University Press, 2011, p. 257.

³³ Resultan interesantes al respecto del contexto y la procedencia las reflexiones de Trevor OWENS: «Digital Sources & Digital Archives: The Evidentiary Basis of Digital History (Draft)», en *Trevor Owens. Digital Centered History*, 2015, disponible en <http://www.trevorowens.org> (consultado el 20 de enero de 2017).

³⁴ Toni WELLER (ed.): *History in the Digital Age*, Londres-Nueva York, Routledge, 2013, p. 212.

³⁵ Sirvan como ejemplo Pares. Portal de Archivos Españoles e Hispana, Apenet y Europeana.

bles para los investigadores. Un buen ejemplo nos los proporciona el proyecto GDELT³⁶ (Global Database of Events, Language, and Tone), una gigantesca base de datos abierta de la sociedad humana creada a partir del seguimiento de los medios de comunicación mundiales en cada país en más de 100 idiomas. En este contexto, en 2015 concluyó la digitalización de 3,5 millones de libros americanos en inglés, de dominio libre, y su procesado, que están disponibles en Google BigQuery, un programa de *big data* que permite ejecutar búsquedas en SQL, en este caso sobre 215 años de historia, en unos segundos. Además de los metadatos de los libros, se han elaborado listas de nombres, organizaciones, lugares, materias, tono (positivo o negativo), polaridad, porcentaje de pronombres, actividad del estilo... Lo que permite búsquedas complejas sobre el tono, la acción o la neutralidad en los textos.

En la base de este reciente despertar de la historia digital hay algunas iniciativas longevas. Una de ellas es el Inter University Consortium for Political and Social Research³⁷, que existe desde 1962 y está integrado por más de 760 universidades del mundo, siendo un proyecto de larga duración para la conservación de datos y resultados de investigación electrónicos en el área con más de cincuenta años de existencia, que también proporciona a los investigadores sociales formación en métodos cuantitativos de análisis, así como herramientas para su explotación. También destacan otras iniciativas como el Roy Rosenzweig Center for History and New Media³⁸ (George Mason University), un centro multidisciplinar que, desde 1994, desarrolla recursos didácticos en línea, colecciones y exposiciones digitales, *software* de código abierto y capacitación en alfabetización y habilidades digitales. Y aún otros más recientes como el Programming Historian³⁹, que funciona como una revista científica y está centrado en la publicación en abierto de lecciones y tutoriales sobre interfaces de programación de aplicaciones (API) como Zotero, gestión de datos con Markdown y Github, cartografía y sistemas de información geográfica (GIS) con QGIS, análisis de redes, etc., en general basados en herramientas populares de las humani-

³⁶ Véase <http://www.gdeltproject.org/>.

³⁷ Véase <https://www.icpsr.umich.edu/icpsrweb/>.

³⁸ Véase <https://rrchnm.org/>.

³⁹ Véase <http://programminghistorian.org>.

dades digitales como las citadas y otras como MALLETT, Omeka, etc. Sin olvidar History Lab⁴⁰, el autodenominado archivo de la libertad de información, una iniciativa de la Universidad de Columbia que permite el acceso a millones de documentos desclasificados, tratados con herramientas de inteligencia artificial para facilitar su acceso. O, en una escala menor, la recopilación de bases de datos en línea de la AHCESR⁴¹.

Por contra, otros proyectos han quedado en el camino, como los efímeros Digital History Project (digitalhistory.unl.edu/index.php, 2012) o Digital Historians (digitalhistorians.org, 2013). Como es fácil deducir, el espacio donde se mueve la historia digital prioritariamente es el anglosajón, donde la American Historical Association publica desde 2014 una selección de recursos sobre la materia⁴².

¿Tiene futuro la historia digital?

¿Es posible garantizar la conservación permanente de los datos, documentos y publicaciones digitales? Una pregunta tan difícil como arriesgada de contestar, a pesar de lo cual trataremos de ofrecer algunos argumentos por el sí condicional. En el lado negativo de la balanza tenemos la omnipresente obsolescencia, las pérdidas parciales debidas a las tasas de error y abundantes experiencias sobre desastres informáticos que han concluido con la pérdida total o parcial de valiosas fuentes de información. Factor importante es también el ciclo de vida de empresas y productos, que concluyen en la desaparición sin dejar rastro. En fin, la acción humana es otro elemento a añadir a este lado del fiel, pero aun y siendo un tema de innegable interés no entra en el ámbito de este análisis.

⁴⁰ Véase <http://history-lab.org>.

⁴¹ *Association des Historiens Contemporanéistes de l'Enseignement Supérieur et de la Recherche*, disponible en <http://ahcesr.hypotheses.org/bases-de-donnees>.

⁴² Véase <https://www.historians.org/teaching-and-learning/digital-history-resources/resources-for-getting-started-in-digital-history> (consultado el 2 de febrero de 2017). También cabe mencionar la selección de investigaciones y proyectos norteamericanos que cada año, desde 2013, publica Rebecca Onion en la revista *Slate*, disponible en <http://www.slate.com/> (consultado el 2 de febrero de 2017). En Europa los proyectos se visibilizan en el entorno de las humanidades digitales, como la Asociación Europea de Humanidades Digitales (<http://eadh.org/>).

En el lado positivo están los estándares que llevan años de existencia (JPEG, TIFF, PDF) y tantos otros que desde hace décadas permiten guardar y recuperar imagen y texto, y todo indica que lo serán por mucho tiempo. También los esfuerzos por lograr sistemas estables, soportes imperecederos, además de la tecnología web que facilita el intercambio, la ubicuidad e incluso el rescate *in extremis*, como viene demostrando desde su fundación Internet Archive.

Por otra parte, conviene ser realistas respecto de las fuentes documentales que hemos heredado hasta ahora. Siempre se hace referencia a los soportes estables del pasado, pero olvidamos que, junto con la piedra y las tablillas de barro, muchos documentos solemnes se hacían sobre tejidos (lino, seda...); que junto con los jeroglíficos en paredes se usaba el papiro; que junto con el pergamino se utilizaba la corteza de árboles, la madera, la pizarra y, sobre todo, la cera para escribir, y no queda nada, sin olvidar los soportes metálicos (bronce, cobre y metales preciosos) reutilizados. Junto con los libros impresos y los códices encuadernados, la mayoría de la producción era a base de documentos sueltos, hojas impresas... mucho más fáciles de perder y destruir.

A pesar de la mucha incertidumbre, parece razonable que en el transcurso de los próximos años la permanencia y la accesibilidad lleguen a solucionarse. Y que se alcance algún tipo de convención que equilibre los derechos económicos de la propiedad intelectual con el derecho al conocimiento y a la conservación del patrimonio histórico digital. Pero ambas cuestiones continúan siendo un escollo insalvable, y si la primera es de orden tecnológico, la segunda es legal, y nos sitúa una vez más ante la paradoja tecnológica: retos globales que requieren soluciones globales, mientras las herramientas, las leyes, son nacionales. Si bien, a falta de una respuesta global a la gestión de los derechos, está la acción social como las iniciativas Open Acces⁴³ (contenidos) y *Open Source*⁴⁴ (sistemas). Una alternativa.

Con todo, aunque las brumas se despejen y la edad oscura sea un mal sueño digital, nos enfrentamos al menos a otros dos retos vinculados a la cantidad y al tiempo. El primero escapa a la escala

⁴³ Véase <https://www.plos.org/open-access> y <http://www.budapestopenaccessinitiative.org/>.

⁴⁴ Véase <https://opensource.org/>.

humana: la sobreabundancia de información, más allá de lo sostenible por sus costes y utilidad. El segundo nos sitúa ante la posibilidad de que la información sea total o parcialmente inaccesible, incluso antes de que transcurra el plazo legal para su consulta, y eso allá donde la legislación establezca plazos precisos. En España, al menos, se añade el desconocimiento acerca de la cantidad y la accesibilidad de los documentos electrónicos producidos hasta ahora por las Administraciones Públicas y poder medir el alcance de nuestra amnesia digital.

En conclusión, es difícil ofrecer certezas acerca de este reto amplio y complejo más allá de lo que lo hiciera Jeff Rothenberg⁴⁵, el autor con el que hemos abierto este artículo: «Digital documents last forever-or five years, whichever comes first».

⁴⁵ Jeff ROTHENBERG: *Avoiding Technological Quicksand: Finding a Viable Technological Foundation for Digital Preservation. A Report to the Council on Library and Information Resources*, Washington DC, CLIR, 1999.

109 ayer



Marcial
Pons