



MICHAEL MOSS,  
DAVID THOMAS,  
TIMOTHY GOLLINS

## Fibras artificiales

Las implicaciones de lo digital  
para el acceso archivístico<sup>1</sup>

### Artificial Fibers

The Implications of the Digital  
for Archival Access

Michael Moss, [michael.moss@northumbria.ac.uk](mailto:michael.moss@northumbria.ac.uk)  
I-School, Northumbria University, Newcastle upon Tyne  
David Thomas  
I-School, Northumbria University, Newcastle upon Tyne  
Timothy Gollins  
University of Glasgow, Glasgow

Citación: Moss M., Thomas, D. y Gollins T. (2020). "Fibras artificiales. Las implicaciones de lo digital para el acceso archivístico". *Tábula*, n. 23, pp. 55-75

Traducción: Alejandro Delgado y Luis Hernández Olivera  
Versión original: "Artificial Fibers - The Implications of the Digital for Archival Access", *Frontiers in Digital Humanities*. 5:20. doi: 10.3389/fdigh.2018.00020. Traducido y publicado con la autorización de los autores y de la revista  
Recibido: 5-8-2020. Aceptado: 23-12-2020  
DOI: <https://doi.org/10.51598/tab.842>

## Resumen analítico / Analytic summary

Este trabajo explora la forma en que los actuales métodos y enfoques archivísticos están seriamente cuestionados por los desafíos que provoca la transformación digital. La disponibilidad de los documentos digitales ha supuesto que a los usuarios se le planteen nuevas necesidades y se abran nuevas posibilidades, incluidos los nuevos modos de consultar. La naturaleza de los mismos archivos está cambiando —están pasando de ser colecciones de textos individuales a ser datos minuciosamente estudiados para darles sentido. Han aparecido nuevas herramientas y técnicas que ya están disponibles y que ofrecen nuevas y radicales posibilidades para la investigación, pero esto comporta nuevos retos en cuanto a la confianza y al gran volumen de documentos a tratar. Los enfoques tradicionales que atribuyen metadatos para facilitar la búsqueda del material relevante y la consideración de los documentos digitales como algo similar al papel electrónico no son viables. Lo que se necesita es una nueva aproximación en la que los archiveros y los investigadores vean los archivos como colecciones de datos que son susceptibles de análisis usando toda una gama de sofisticadas herramientas y que son capaces de ser interpretados con toda una gama de fórmulas diferentes.

CORREO ELECTRÓNICO | CONSERVACIÓN | ACCESO A LOS DOCUMENTOS | ARCHIVOS | METADATOS | PROVISIÓN DE SENTIDO | VALORACIÓN

This article explores how current methods and approaches in archives are under serious challenge because of the changes brought about by the move to the digital. The availability of digital records has meant that new needs and new possibilities have opened up for users, including new ways of reading. The nature of archives themselves are changing—they are moving from being collections of individual texts to be pored over to data to be made sense of. New tools and techniques have emerged and are available now which offer radical new possibilities for research, but these bring new challenges about trust and the sheer volume of records to be handled. The traditional approaches of applying metadata to facilitate the finding of relevant material and of regarding digital documents as something like electronic paper is no long viable. What is needed is a new approach in which archivists and scholarly researchers see archives as collections of data which are capable of analysis by a range of sophisticated tools and which are capable of being interpreted in a range of different ways.

EMAIL | PRESERVATION | ACCESS RECORDS | ARCHIVE | METADATA | SENSE-MAKING | APPRAISAL

## El reto de lo nuevo

La práctica archivística sigue anclada en los procesos artesanales. Con solo un vistazo a la colección aleatoria de miles de correos electrónicos mal ordenados, que se encuentran en Wiki-Leaks, queda claro que el acceso al contenido digital nacido no puede continuar proporcionándose a través de los catálogos convencionales. De hecho, al tratar sus propias colecciones de lo que la United Kingdom Foreign and Commonwealth Office (FCO) todavía llama telegramas nocturnos, la FCO está experimentando con nuevas herramientas de provisión de sentido para ayudar a las embajadas a interpretar los flujos de datos que reciben (Greenhalgh, 2014). Estas ideas y sus aplicaciones son herederas de las iniciativas para comprender los mensajes de las operaciones militares emitidos a finales del siglo XX por los EE.UU. (Grishman y Sundheim, 1996). Junto con los incluso más novedosos modos de comunicación digital, como el chat de texto interactivo (por ejemplo WhatsApp) y los crecientes usos de plataformas de medios sociales tanto públicas como semi-privadas (por ejemplo, el uso de Twitter por los departamentos de Whitehall<sup>2</sup> y el intento de los gobiernos escoceses de uso interno de Yammer), significa que los documentos a los que el archivo se está confrontando son grandes acumulaciones de ‘materiales’.

Es ilusorio imaginar que el orden puede imponerse a todo menos a una fracción del contenido, ni siquiera en el momento de su creación. Incluso para los documentos digitales ‘convencionales’, como sabemos, al menos en la administración pública del Reino Unido, los registros y ficheros han casi desaparecido (Allan, 2014, 2015). Los correos electrónicos de Enron que se pusieron a disposición de la investigación judicial eran 620.000 correos variados y el único modo en que se

pudo darles sentido fue usando técnicas computacionales y estadísticas avanzadas del Language Technology Institute at Carnegie Mellon University (Klimt y Yang, 2004b) y del Departamento de Informática de la Universidad de Columbia (Prabhakaran et al., 2014). El uso de tales técnicas se está generalizando en la comunidad forense digital (con herramientas comerciales como Nuix)<sup>3</sup>. Su potencial de aplicación a materiales archivísticos podría decirse que solo sería una sofisticada ampliación al mundo digital del concepto fundamental de catalogación (TNA, 2016). La catalogación siempre ha tenido que ver con la provisión de sentido post-hoc, incluso aunque a veces estuviera abierta a acusaciones de ‘manipulación histórica fortuita’ (Lynch, 2003, 196).

### Las necesidades cambiantes de los usuarios del archivo

La comunidad archivística ha tardado en reconocer el reto del acceso al contenido digital, quizá pensando ingenuamente que la aleatoriedad de los actuales motores de búsqueda serviría. Gran parte de la comunidad no se ha dado cuenta de que los usuarios no solo necesitarán, sino que serán capaces de, implementar sofisticadas herramientas y servicios que permitirán que el contenido digital sea interpretado de maneras radicalmente diferentes como en el FCO; o incluso históricamente como en la red de relaciones de Francis Bacon, un proyecto de la Universidad Carnegie Mellon (Rea, 2015). La comunidad archivística tiene que dejar de preocuparse por los incómodos metadatos al tratar con objetos digitales y comprometerse con las comunidades de estadísticos, matemáticos e informáticos. Tales equipos, en Carnegie Mellon y en Columbia, están desarrollando ya nuevas herramientas y servicios de provisión de sentido que no imponen cargas innecesarias a los creadores de contenido.

Este enfoque transdisciplinar debe basarse en la interacción continuada entre la oferta y la demanda; entre, por una parte, las preocupaciones profesionales y los problemas prácticos a los que se enfrentan los archiveros que intentan preservar documentos digitales; y, por otra, las cambiantes necesidades de los investigadores. Durante mucho tiempo los archiveros se han concentrado en el lado de la oferta y han descuidado el de la demanda. Esto debe cambiar. Este enfoque sobre la oferta está bien articulado en un trabajo de Clifford Lynch en el que afirma: ‘deberíamos evitar enfatizar demasiado las *nociones preconcebidas* [el realce es nuestro] acerca de las comunidades de usuarios cuando se crean colecciones digitales, al menos en parte porque somos bastante malos identificando o prediciendo estas comunidades-objetivo’ (Lynch, 2003, p. 196). Incluso en el ahora maduro campo de la conservación digital el frecuentemente citado modelo de referencia ‘OAIIS’ (Lavoie, 2014) contiene el concepto de ‘Comunidad Designada’ para proporcionar la base de una justificación (término nuestro) de las

costosas acciones emprendidas para conservar una colección. Tales nociones preconcebidas confirman que el valor y la utilidad de una colección son intrínsecamente problemáticos y deben implementarse, desde nuestro punto de vista, con mucho cuidado.

Más recientemente, Tom Schofield y otros trabajaron para entender a los usuarios del archivo de la editorial de poesía Bloodaxe Books. Su investigación inicial mostró que ‘prácticamente todos los aspectos identificados como interesantes por los participantes no iban a describirse en los metadatos del próximo catálogo y, por lo tanto, no estarían representados en los futuros interfaces del archivo (Schofield et al., 2015). Ya en 1987, Bruce Dearstyne, escribía en *The American Archivist* (Dearstyne, 1987) y expresaba críticas muy similares. Citaba a Roy Turnbaugh (Turnbaugh, 1983, p. 451), quien había sugerido que ‘los archiveros producen instrumentos de descripción que son o ignorados o difíciles de utilizar y que se aferran a conceptos obsoletos inadecuados para los enfoques y necesidades de los investigadores modernos’.

Creemos, como Lynch y Dearstyne, que el énfasis en las ‘nociones preconcebidas’ de las potenciales comunidades de usuarios es realmente inútil. Pero lo más importante es que esto no supone un abandono intelectual del lado de la demanda por parte de los archiveros. Muy por el contrario, la emergencia en otros campos de las ciencias de la información de técnicas que pueden ‘proporcionar sentido’ ‘a la demanda’ de colecciones indiferenciadas de información proporciona una oportunidad para una transformación de la práctica archivística. En Suiza, por ejemplo, Basma Makhlof Shabou y Maria Sokhn (Shabou y Sokhn, 2017, p. 219) están trabajando para ‘poner en valor el patrimonio histórico mediante una plataforma de diseño centrado en el ciudadano’ llamada City-Zen. La idea es proporcionar al turista utilidades que den sentido a los datos disponibles en una diversidad de formatos y de localizaciones.

Estas herramientas emergentes eliminarán la necesidad de enfatizar e imponer una sola estructura que, por su misma naturaleza, estará sesgada (Pitti, 2006). Estas herramientas permiten que cada usuario determine su propia visión del archivo o colección de datos en sus propios términos. Por lo tanto, la comprensión de este nuevo enfoque del lado de la demanda, se vuelve crucial para comprender lo que el archivo tiene que contener para dar soporte a esta comunidad emergente de nuevos usuarios y a sus demandas. Estas nuevas demandas pueden, a su vez, reforzar una nueva visión de la oferta y, casualmente, las mismas herramientas que los usuarios desean aplicar a las colecciones serán demandadas por los archiveros para comprender el nuevo paisaje de los documentos digitales que encuentran en el lado de la oferta.

Al cambiar su perspectiva epistemológica, estos nuevos e imaginativos usuarios de los archivos y de recursos relacionados están comenzando a decir que usando tales técnicas podemos hacer cosas excitantes con sus datos. Por ejemplo,

podrían visualizar los asientos del catálogo usando NGram, o coger una lista detallada de correspondencia y visualizar los enlaces, o como en el ejemplo suizo citado más arriba explorar el patrimonio cultural de una ciudad. Se han acometido muchas investigaciones para visualizar documentos y los complejos datos generados por estadísticas y el análisis matemático del contenido (Ahlberg y Shneiderman, 1994). Todo lo que tienes que hacer es explorar las opciones que se te ofrecen para ver tu galería de fotografías en tu i-Phone. Como apunta Kalpesh Padia: ‘Muchos archivos web, como Archive-It<sup>4</sup>, California Digital Library<sup>5</sup>, Library of Congress<sup>6</sup> y Pandora - Australia’s Web Archive (Pandora)<sup>7</sup> proporcionan un interfaz textual para interactuar con las colecciones archivadas (Padia, 2012, p. 17). Además, un proyecto del MIT desarrolló un sistema ‘Themail’ para visualizar los contenidos de los buzones de correo. Los datos que Themail visualiza consisten en los ficheros procesados de los buzones de correo. Themail comienza con un fichero de correo en forma de uno o más ficheros mbox, que se procesan después aplicando un algoritmo de puntuación de palabras clave. Esta aplicación da como resultado un fichero de datos que puede consultarse a través de la visualización (Viégas et al., 2006; Themail)<sup>8</sup>. Estos desarrollos inevitablemente darán como resultado una reconfiguración de la práctica y tendrán ramificaciones sobre las formas convencionales de hacer las cosas, a medida que las nuevas herramientas y servicios estén disponibles comercialmente.

Los archiveros no han reconocido el impacto que el paso hacia lo digital ha tenido sobre los usuarios porque se han centrado más en el lado de la oferta que sobre el de la demanda. Sin embargo, los investigadores, a menudo al margen de las principales corrientes archivísticas, están desarrollando nuevas herramientas para permitir la búsqueda individualizada de recursos archivísticos. Un aspecto significativo de este movimiento hacia lo digital es que ha llevado a un nuevo modo de consultar los archivos que describimos a continuación.

## Los modos cambiantes de consultar los archivos

Hasta hace poco, la historia ha estado basada en gran medida en textos y ha confiado en los libros y documentos que proporcionaban los archivos. Algunos autores, en particular Tim Hitchcock (Hitchcock, 2015), han sugerido que lo digital permite que la historia se amplíe para abarcar otras fuentes, incluidos el sonido, el video e incluso lo táctil. En efecto, algunos historiadores se han basado en la historia oral o en las tradiciones indígenas. En este artículo, nos centramos en lo textual que, creemos, todavía constituye la base de la investigación histórica actual.

En 1938, Cleanth Brooks y Robert Penn Warren publicaron su seminal *Understanding Poetry* (Brooks y Warren, 1938), que determinó el análisis de las

obras literarias de los siguientes 60 ó 70 años. Durante ese período, el método predominante para el análisis de las obras literarias fue lo que se llama ‘lectura detallada’, un estudio minucioso de los textos que se enfoca sobre la obra de arte como un objeto autónomo que puede ser analizado en sus propios términos (Davis, 2011). Esta literatura de la ‘lectura detallada’ se ha expandido desde los estudios literarios a los de humanidades de forma general. En 2000, Franco Moretti, un erudito italiano, comenzó a interesarse por la literatura mundial. Al darse cuenta de que no hay posibilidad de consultar más que una minúscula proporción de la literatura mundial, propuso un nuevo enfoque que llamó ‘lectura a distancia’, en el que, en lugar de estudiar los textos, se buscan temas en la historia literaria (Moretti, 2000). Desde entonces, Moretti (2013) se ha preocupado más por lo digital y sus puntos de vista se han extendido hacia las humanidades.

Los historiadores e investigadores de las humanidades digitales William Turkel, Kevin Kee, y Spencer Roberts (Turkel et al., 2013, p. 62) han argumentado que la lectura atenta de los textos es una tarea imposible en la era digital. Citan a Cohen (2011), que señaló que, aunque un solo historiador pudiera haber sido capaz de consultar los 40.000 memorandos publicados en la Casa Blanca por la administración Johnson, evidentemente no hubiera podido tratar los cuatro millones de correos electrónicos enviados cuando Clinton estuvo en el cargo. En el futuro, los usuarios tendrán que confiar en sofisticadas herramientas analíticas para permitir la lectura a distancia de un gran volumen de material.

Tales herramientas no deberían confundirse con una obsesión por la ‘búsqueda’ que solo permite encontrar lo esperado por el investigador. En el contexto de internet, las ideas preconcebidas de los usuarios, no importa cuán extremas o abstrusas sean, siempre pueden justificarse y reforzarse en algún punto de la enormidad del ese gigantesco recurso, los resultados contrarios a la evidencia más grande, simplemente no se ven. No debemos permitir que este modo de interacción sea el único que ofrezcamos a nuestras colecciones. Deberíamos, por supuesto, dar soporte a la búsqueda, pero proporcionando otras herramientas analíticas que puedan animar a la adopción de una visión más amplia, y al descubrimiento de patrones y comprensiones actualmente ocultas a simple vista. Esto es lo que el ya citado City-Zen pretende hacer.

David Weinberger, en *Too Big to Know*, caracteriza esto como ‘lectura de formato amplio’, que según algunos comentaristas ‘permite y fomenta el pensamiento de formato amplio’ (Weinberger, 2011, p. 99). El autor sugiere que ni el formato amplio ni la lectura detallada eran ‘una buena opción para estructurar el mundo. Quizá las redes entrelazadas reflejan el mundo de manera más precisa’ (Weinberger, 2011, p. 115). Identifica cinco propiedades del mundo en red que encontramos en nuestros navegadores: ‘abundancia, enlaces, libre de permisos, público e irresoluto’, que tomadas en conjunto niegan la práctica tradicional (Weinberger, 2011, p. 174).

Esto requiere que los usuarios confíen en las herramientas de provisión de sentido y de descubrimiento que hacen posible trabajar de nuevas maneras. Tim Hitchcock, en su blog *Historyonics*, habla de que los historiadores tienen que desarrollar un macroscopio metafórico, un dispositivo que haga posible ver tanto objetos grandes como pequeños al mismo tiempo (Hitchcock, 2014). El autor cita el *History Manifesto* de Jo Guldi y David Armitage (Guldi y Armitage, 2014), que argumenta que una vez armados con un 'macroscopio'... los historiadores deberían continuar con un análisis acerca del modo en que podría usarse el 'big data' para renegociar el rol del historiador y de las humanidades en general (Hitchcock, 2014). Los autores de *Exploring Big Historical Data: The Historian's Macroscope*, Shawn Graham, Ian Milligan y Scott Weingart argumentan además:

*'No estamos dando a entender que ésta será la forma en la que los historiadores 'harán' historia cuando llegue el big data; más bien, es solo una pieza de la caja de herramientas, una manera más de tratar con las 'grandes' cantidades de datos con las que los historiadores están teniendo que lidiar ahora. Es más, un 'macroscopio', una herramienta para observar lo muy grande, sugiere deliberadamente la mesa de trabajo de los científicos, en la que el investigador se mueve entre diferentes herramientas para explorar diferentes escalas, tomando notas en el cuaderno de investigación (Graham et al., 2015, p. xvi).*

El movimiento hacia lo digital ha supuesto un cambio en el modo en el que los archivos se consultan. A causa de la vasta escala de los recursos disponibles, los investigadores se están moviendo ahora de la lectura detallada de los documentos a la lectura a distancia, desde un microscopio a un macroscopio. Al mismo tiempo, toda la naturaleza del archivo en sí mismo está cambiando.

## La naturaleza cambiante del archivo

Estos procesos están cambiando la naturaleza del archivo, que está pasando a una reconceptualización como datos a los que proporcionar sentido. Esto no es solo para el documento oficial, sino para toda la masa de aportaciones de los medios sociales y comentarios de los usuarios (Merrin, 2014, p. 152). Sin duda alguna, el archivo como un todo, siempre ha sido difícil de comprender. De hecho, el acceso a los documentos asociados, en el mundo en papel, no resultaba tan fluido como internet lo ha hecho en el mundo digital. Por ejemplo, se podían encontrar noticias en los periódicos, documentos privados en archivos personales en las bibliotecas, y archivos y documentos familiares en empresas comerciales, etc.

Además, el archivo ya no es estático. Los patrones de bits no solo son intrínsecamente difíciles de autenticar (Allison et al., 2010), sino que al archivo se le añaden constantemente comentarios de los usuarios, se recatologa y copia, y se

pone a disposición en diferentes localizaciones, que usualmente son públicas e 'irresolutas'. Un ejemplo de este input y compromiso públicos es el recientemente lanzado por la National Archives and Records Administration: "History Hub - A support community for history enthusiasts, researchers, citizen archivists, family historians, archival professionals and open government advocates" (History Hub, 2015). Michelle Caswell propone un argumento similar en un reciente debate: 'los archiveros deberían invitar a los usuarios, así como a otros actores externos al proceso archivístico, a participar en la descripción archivística utilizando lenguajes, categorías, sistemas y normas que sean significativos para ellos' (Caswell, 2016).

El Departamento de Estado de los EE.UU. tiene un gran archivo de telegramas, pero también tienen copias de este archivo Wiki-Leaks y las personas que se lo han descargado y analizado. Al hacerlo, algunas consignas largo tiempo apreciadas por los archiveros, como las jerarquías, el orden original y la procedencia, se vuelven imposibles de establecer con autoridad. Muy poco material textual nacido digital ha entrado en el dominio público por los debidos procesos legales de cualquier lugar del mundo, con la notable excepción de la legislación del Reino Unido (<http://www.legislation.gov.uk/>), con sentencias de su Tribunal Supremo (<https://www.supremecourt.uk/news/latest-judgments.html>) y tribunales menores (<https://www.judiciary.gov.uk/judgments/>), ahora públicamente disponibles en línea. Además, hasta donde sabemos, el corpus de Enron sigue siendo la única colección a gran escala de correos electrónicos autenticados disponibles para los investigadores. Hay una gran cantidad de corpus filtrados en WikiLeaks, pero estos no han sido autenticados de la misma manera que Enron (Klimt y Yang, 2004a, b). Por tanto, es muy difícil para la comunidad archivística estimar cuán dramática será la reforma de la práctica, o cuáles serán las necesidades del usuario cuando implementen nuevas herramientas analíticas. Sin embargo, cualquiera que use regularmente grandes colecciones de material digitalizado, como Ancestry.com, Trove en Australia, British Newspapers en línea, o Google Books, reconocerá casi instintivamente a qué se refiere el mencionado David Weinberger al proponer redes entrelazadas. Las páginas web de historia familiar, como Ancestry.com, ya ofrecen a los usuarios utilidades que enlazan los datos que se encuentran en las diversas colecciones sobre las que tienen derechos, como censos, registros de nacimientos, matrimonios y defunciones, y periódicos.

El archivo no solo se está consultando de nuevas formas, y su misma naturaleza está cambiando desde los textos a los datos que se van a analizar, sino que como mostramos en la siguiente sección, el movimiento hacia lo digital ha hecho posible toda una gama de maneras no textuales de experimentar los archivos.

## Nuevos modos de experimentar el archivo

Un reciente informe de Ian Chowcat al JISC (anteriormente el Joint Information System Committee), que da soporte a la educación de bachillerato y superior en el Reino Unido, indica que la generación de jóvenes que entrarán en las universidades a partir del 2020 utilizarán interfaces basadas en el tacto o el gesto. Verán experiencias en línea y fuera de línea combinadas perfectamente y, en consecuencia, parecerán tener altas preferencias visuales (Chowcat, 2015, p. 5). Esto nos da algunas claves en cuanto al camino a seguir al pensar el diseño de interfaces y la prestación de servicios.

Sin embargo, es importante no ver esto como algo que va a suceder; ya está ocurriendo ahora. Si no lo ha hecho, vaya a la página web de la Virtual St. Paul Cross, que permite experimentar el sermón de John Donnes ‘El día de la pólvora’ de 1622 (Proyecto Virtual St. Paul’s Cathedral)<sup>9</sup>. El sitio combina el texto del sermón con el uso de software de modelado arquitectónico y software de simulación acústica, de modo que usted está allí, en ese frío día de noviembre del reinado de Jacobo I, cuando Donne pronunció su sermón. Éste no es un ejemplo aislado. Hay otros muchos ejemplos que reflejan el modo en que lo digital confunde la temporalidad. Sin embargo, como Dave Nicholas ha mostrado, éste es el modo en que los usuarios contemporáneos trabajan. Saltan de una cosa a otra en tiempo real (Nicholas, 2007, p. 125). Pueden observar a John Donne durante un tiempo, pero luego saltan a Tim Hitchcock, que está trabajando para recrear la ambientación sonora de la sala del tribunal de Old Bailey. Esta ambientación de sonido recrea la experiencia auditiva del acusado –cómo se sentía al hablar al poder, y qué percibía cuando el poder le hablaba desde el estrado (las Voces de Old Bailey)<sup>10</sup>. El proyecto Virtual St. Paul’s Cross y la ambientación sonora de Hitchcock nos llevan más allá de la textualidad del archivo hacia algo que se aproxima a la experiencia original de la audiencia del sermón de Donnes o del juez, el jurado, los abogados y los acusados en el tribunal. Como Holger Schott Syme ha argumentado, las declaraciones de los testigos que sobreviven en la forma de testimonios escritos y que son una característica fundamental de los documentos textuales de los tribunales fueron leídos en el tribunal por un escribano (Schott Syme, 2003, p. 109). Es importante, como él ha enfatizado, que fue esa lectura en voz alta la que constituyó la evidencia, y no el texto escrito. A esto podríamos añadir las colecciones de sermones pronunciados desde los púlpitos. El usuario solo será plenamente consciente del propósito de estos sitios si se le puede animar a extender el “tiempo de permanencia” en ellos. Esto puede ser, por ejemplo, y siguiendo el modelo City-Zen, caminar entre Old Bailey y St. Paul.

El paso a lo digital que ha cambiado la naturaleza del archivo y permitido nuevas formas de consulta también ha posibilitado las experiencias visuales y acústicas, y si creemos a Tim Hitchcock, también tangibles. Estos cambios radicales abren nuevas e interesantes posibilidades a la investigación.

## Nuevas posibilidades para la investigación

Estas nuevas tecnologías hacen posible realizar alguna investigación notable. Tomemos el ejemplo del trabajo sobre los correos electrónicos de Enron en Columbia. Oweb Rambow, uno de los investigadores, escribió a los autores:

*“Vemos profundas diferencias entre el uso del lenguaje por parte de las personas con poder y las personas sin poder, lo que nos permite predecir quién tiene poder en un diálogo. Nos hemos preguntado cómo cambia este comportamiento relacionado con el poder cuando incorporamos el género de los participantes al análisis del discurso. Hemos encontrado profundas diferencias en el uso del lenguaje entre hombres y mujeres con poder, y también entre entornos de género femenino y no femenino (el entorno de género refleja el género de todos los participantes en el discurso). Estamos investigando el modo en que las redes sociales que pre-existen a un diálogo en particular se conectan con las relaciones de poder”.* (Rambow, comunicación personal).

Éste es precisamente el tipo de técnicas que defienden Guldi y Armitage en el History Manifesto (Guldi y Armitage, 2014). Aunque existen obstáculos para acceder y manipular el contenido nacido digital, particularmente sensibilidades éticas, especialmente la protección de datos y la propiedad intelectual, éstos no deberían obstaculizar la experimentación con el creciente cuerpo de contenido abierto a la exploración digital que está incorporándose al dominio público. El proyecto Digital Panopticon está analizando lo que las técnicas de visualización pueden revelar acerca de la forma general y de los patrones distintivos de los datos, y lo que estos revelan acerca de los diversos procesos mediante los cuales se crearon los datos, y sus restricciones y limitaciones (Digital Panopticon)<sup>11</sup>. Como ocurre con todos los desarrollos técnicos, algunas técnicas tendrán poca utilidad, pero otras permitirán interpretaciones novedosas, por ejemplo la visualización de colaboradores de Wikipedia que confirmaron la existencia de wikipedistas (Zachte, 2011). Sin embargo, esto solo sucederá si hay un diálogo entre las disciplinas y en el archivo entre el lado de la oferta y el de la demanda.

Hasta ahora, hemos presentado una imagen optimista de una revolución digital que ha cambiado la naturaleza del archivo y que invita a nuevos modos de consulta, y de visualizar los archivos, y que facilita nuevos métodos de investigación. Sin embargo, hay que considerar dos cuestiones: la confianza y el volumen.

## Cuestiones de confianza

Uno de los grandes desafíos a los que se enfrentan los desarrolladores de estos nuevos enfoques de la información es la necesidad de asegurar la confianza de los usuarios. Tradicionalmente, los usuarios han tenido que confiar en las habilidades, la honestidad y la amplitud de visión de los catalogadores archivísticos para que les proporcionen la seguridad de que todo el material de potencial interés para ellos ha sido descrito, pero, como hemos mostrado anteriormente, esa confianza se ha perdido. Ahora hay otra cuestión de confianza: los usuarios de sistemas en línea, ya sean las comunes herramientas de búsqueda o aplicaciones de búsqueda mucho más específicas, tienen que confiar en la tecnología.

El usuario está, en un sentido muy real, cambiando la confianza en los catalogadores por la confianza en los tecnólogos. Además, la confiabilidad de algunos motores de búsqueda y webs de redes sociales ha sido cuestionada. Por ejemplo, en 2015, investigadores de la Harvard Business School, la Columbia Law School y Yelp argumentaban que “Al mostrar de forma destacada el contenido de Google en la respuesta a las peticiones de búsqueda, Google puede utilizar su posición de dominio en la búsqueda para ganar clientes para su contenido”. (Luca et al., 2015, p. 1).

## Cuestiones de volumen

La cuestión fundamental a la que nos enfrentamos es la del volumen. Tanto si los gestores de documentos lo capturan todo como si están comprometidos en un proceso de selección, la consecuencia inevitable de lo digital será que ingresaremos muchos más documentos, ya sea para satisfacer la demanda o porque tratar de desenredar los conjuntos de correos electrónicos y otros conjuntos de datos digitales es muy difícil. Klimt y Yang solo lograron reducir los correos electrónicos de Enron en dos tercios después de haber limpiado un conjunto de aproximadamente 600.000 correos electrónicos dejándolo en 200.000 (Klimt y Yang, 2004a, p. 1). Ésta es una magnitud menos efectiva que las actuales prácticas de valoración. Esto se debe a la rápida reducción en el coste de creación y almacenamiento de datos, paralela a una explosión de usuarios de internet de 2 billones en 2010 a 4 billones en 2017 (Internet World Stats, 2017). En el Reino Unido, la Army Historical Branch ha recibido desde 2002 10 millones de documentos ‘declarados’ y 60 millones de ‘no declarados’ (Evans, 2015). Viktor Mayer-Schönberger cuenta la historia de su padre que, siendo un adolescente, recibió de regalo una cámara Kodak Brownie en los años treinta, pero se le advirtió de que las fotografías eran caras y solo debería utilizarla en ocasiones especiales. Como resultado, en los años siguientes solo sacó tres docenas de fotografías de eventos familiares importantes y de las montañas que escaló (Mayer-Schönberger, 2011, p. 45).

Ahora las cámaras vienen incluidas en los teléfonos, sin costes de procesamiento ni de impresión, con dispositivos de almacenamiento portátiles cada vez más grandes. La última tarjeta de memoria SanDisk para smartphones almacenará 36.000 fotografías –suficiente incluso para el turista con palo selfie. Además, los consumidores tienen a su disposición almacenamiento gratuito en toda una gama de sitios –Dropbox, Box, Google Cloud, etc. Google comenzó a ofrecer una cuenta gratuita de correo electrónico con un gigabyte de capacidad el 1 de abril de 2004, con el fin de ‘almacenar gratuitamente para que nunca se tuviese que volver a borrar un correo’. Recientemente, han publicitado su teléfono Google Pixel 2, con ‘almacenamiento ilimitado’, a pesar de la desafiante publicidad en contra (Smith, 2017). La mayoría de los nuevos PCs vienen con un terabyte de almacenamiento, algo inimaginable hace solo una década. ¿Cuánto pasará antes de que un exabyte se convierta en la norma?

Las empresas y los gobiernos tienen acceso a un almacenamiento propio cada vez más barato y a soluciones basadas en la nube de alta calidad. De conformidad con la ley de Kryder (el equivalente en almacenamiento a la ley de Moore para la capacidad de procesamiento informático), la capacidad y el coste del espacio en disco duro se reducen a la mitad cada 18-24 meses (Walter, 2005, pp. 32-33). Sin embargo, esto está lejos de aceptarse universalmente; el trabajo reciente de una de las más influyentes figuras en el campo de la conservación digital, David Rosenthal, sugiere que tales ‘leyes’ puede que no sean universales y que la expansión del espacio de almacenamiento alcance límites físicos en un futuro previsible (Rosenthal, 2017).

Además, gran parte del material que está almacenado consiste en efímera y duplicados –copias de correos electrónicos, fotografías y otros documentos que se almacenan automáticamente en múltiples copias y en numerosos dispositivos portátiles, equipos de escritorio, servidores y sistemas de copias de seguridad. Aunque el coste del almacenamiento ha estado bajando, no puede decirse lo mismo de los costes de gestión de documentos. En consecuencia, el coste de ‘olvidar’ los datos mediante la eliminación selectiva requiere más esfuerzo y es más costoso que conservarlos. Esto inclina enfáticamente la situación por defecto hacia la conservación o quizá de manera más precisa hacia ‘guardar cosas’. Más aún, borrar documentos digitales es difícil. Las copias de los correos electrónicos se almacenan en los ordenadores de los emisores, en los de los receptores o en un servidor central, y puede que también se almacenen en un buffer o almacenamiento temporal. Presionar la tecla de borrado no borra necesariamente en todas las copias almacenadas o ‘conservadas’.

Los problemas graves surgen cuando los seres humanos necesitan comprender el significado de la información a saco, donde la cantidad de datos que proporciona esta información es demasiado para que una persona la comprenda o demasiado grande para que los recursos humanos disponibles la puedan leer y

analizar. Una cuestión determinada a este respecto es que dado que un conjunto de datos grande contiene muchos millones de palabras, es probable que una búsqueda a texto completo devuelva, al menos, unos pocos ejemplos de lo que se quiera buscar. Así, la hipótesis inicial siempre se refuerza. Éste es un problema que Ted Underwood ha definido como el ‘sesgo de la confirmación’ (Underwood, 2014, p. 66). Underwood da el ejemplo de un investigador que tiene la hipótesis de que la palabra rubor simboliza la consciencia moral en la poesía del siglo diecinueve. El investigador puede ir a una base de datos de fuentes primarias y buscar poemas que contengan tanto ‘rubor’ como ‘consciente’. Si tiene éxito, entonces puede escribir un artículo. Si no, entonces pueden buscarse asociaciones alternativas de palabras – ‘sonrojo’ y ‘vergüenza’, por ejemplo. En una base de datos lo bastante grande, tales enlaces se encontrarán. Underwood sigue afirmando:

*“Es cierto que la búsqueda a texto completo puede confirmar casi cualquier tesis que se plantee, pero puede que esta no sea su característica más peligrosa. El problema más grave es que ordenar las fuentes según su relevancia para la consulta también tiende a filtrar todas las tesis alternativas que no se plantearon. La búsqueda es una forma de minería de datos, pero una forma extrañamente dirigida que solo te muestra lo que ya sabes que puedes esperar”*. (Underwood, 2014, p. 66).

Underwood continúa sugiriendo dos enfoques alternativos. Una es utilizar un algoritmo que busque las palabras que se asocian de manera más comúnmente con “rubor”. Esto da como resultado que esa palabra es ‘ingenuo’, lo cual debilita la idea de que los sonrojos son algo que tiene que ver con la consciencia moral. La otra es utilizar el modelado tópico – que permite que el ordenador organice el lenguaje de una colección en agrupaciones de términos que tienden a aparecer en los mismos contextos (Underwood, 2014, pp. 67–69). Esto permite revelar patrones discursivos que el investigador no estaba necesariamente buscando. Fundamentalmente, todos estos enfoques, desde una búsqueda simple por palabras hasta el modelado tópico, no son el producto inconsciente de una máquina de inteligencia artificial, sino que todos ellos han sido generados por hipótesis humanas.

Aunque lo digital ha abierto nuevas y enormes posibilidades, existen dificultades muy reales – podemos confiar en la tecnología utilizada para buscar tales materiales y cómo podemos abordar los volúmenes gigantescos a los que nos enfrentamos. Un enfoque tradicional entre los archiveros es el de confiar en los metadatos. La validez de este enfoque se discute en el siguiente apartado.

## ¿Metadatos o no?

La doctrina convencional dice que para lograr resultados de búsqueda superiores se necesitan más y mejores ‘metadatos semánticos’ (Duff y van Ballegoie, 2006). Sin embargo, incluso en el mundo analógico, los metadatos no son siempre la respuesta. Por ejemplo, cuando David Thomas, fue responsable de una rueda de prensa del Archivo Nacional (Reino Unido) sobre los expedientes relativos al incendio de Windscale de 1957, la organizó, naturalmente, para que las copias de las descripciones del catálogo que se distribuyeran fueran las más útiles. Al final del evento de prensa, un periodista le dijo que estaba bastante equivocado y que el expediente más interesante estaba en la colección del Northern Production Group de la Atomic Energy Authority (UKAEA) del Reino Unido, titulado de forma poco clara que se describía genéricamente como ‘Documentos de las divisiones de producción de la UK Atomic Energy Authority, que abordan todos los aspectos de la investigación en energía atómica, los procedimientos cotidianos, las funciones administrativas y las relaciones industriales’ (TNA, 2018).

El problema de confiar en los metadatos en el mundo digital es, aparte de los metadatos ambientales que se generan automáticamente, cómo se crean. La solución actual parece estar en que los usuarios sean responsables de idear y aplicar metadatos adicionales, como proporcionar títulos nuevos y más ‘adecuados’ para los correos electrónicos y titular útilmente los ficheros de las colecciones de documentos dentro de sistemas de gestión de documentos electrónicos (EDRMS). Sin embargo, como Michael Moss descubrió, ‘los proyectos que han investigado esto han mostrado que los usuarios solo estarán preparados para adoptar tales convenciones si pueden ver claro el valor añadido para ellos mismos en términos de sus procesos comerciales, lo que incluye el cumplimiento de las regulaciones’ (Moss, 2005, p. 589). Simplemente, no es razonable esperar que los altos directivos actúen como auxiliares de archivo. Esto se ha confirmado recientemente en dos investigaciones sobre la gestión de documentos del gobierno del Reino Unido de Allan (2014, 2015). Los informes señalan:

*Los sistemas existentes que requieren que los usuarios identifiquen documentos que deberían constituir documentos oficiales, y luego los guarden en un EDRMS o en un sistema de archivo corporativo, no han funcionado bien. Los procesos han sido gravosos y el cumplimiento pobre. Como resultado, casi todos los departamentos tienen una gran cantidad de datos digitales almacenados en discos compartidos que están mal organizados e indizados”* (Allan, 2015, p. 1).

Además, en el Reino Unido el Archivo Nacional ha admitido que solo el 33 por ciento de los datos se conservan en EDRMS, el mayor volumen se mantiene desestructurado en discos compartidos (TNA, 2016, p. 10).

Finalmente, para ilustrar los puntos que hemos señalado sobre la naturaleza cambiante del archivo y las problemas que plantea, volvemos a los correos electrónicos que, para la mayoría de nosotros, son la viva encarnación tanto del potencial de los documentos digitales como de sus retos.

## Correos electrónicos

Sin un tratamiento adecuado, los correos electrónicos son invisibles como documentos, pero además, como sabe cualquiera que haya buscado el documento de una decisión en una cadena de correos electrónicos, un correo electrónico tiene poco valor por sí mismo. Solo puede entenderse como parte de una larga y compleja cadena de correspondencia con numerosas pistas falsas. Los hilos de muchos cientos de mensajes de correo dirigidos a decenas de personas pueden contener tanto contexto como contenido que solo se puede obtener visualizando el hilo completo como un gran documento y extrayendo el contexto y el contenido de cientos de crípticos mensajes de correo que individualmente casi no tienen ni contenido ni contexto comprensible.

La filosofía que considera una pequeña proporción de los correos electrónicos como ‘sustantivos’ y valiosos para convertirlos en ‘documentos oficiales’ debe cuestionarse. El enfoque adoptado por los equipos forenses digitales para el correo electrónico debería indicar a los archiveros dónde reside la evidencia real de las transacciones (Waugh, 2014). En el mundo forense, los correos electrónicos se capturan como colecciones completas sin filtrar, lo que enfatiza nuevamente la importancia del lado de la oferta. Solo una vez que se determina el objetivo de una investigación se aplican filtros para limitar los datos a examinar a aquellos que son relevantes (lado de la demanda). El correo electrónico es naturalmente rico en lo que los archiveros describirían tradicionalmente como metadatos ambientales, como el destinatario, la hora y la fecha de envío, etc. Ésta y otras consideraciones han llevado al Archivo Nacional (NARA) de EE.UU. a desarrollar CAPSTONE, que entró en vigor a fines de diciembre de 2016, de conformidad con la Managing Government Records Directive (M-12-18) (Email Management, 2016). Este enfoque sugiere que las agencias del gobierno de los EE.UU. solo deberían capturar documentos para su conservación permanente, a partir de las cuentas de correo de los funcionarios en o próximos de los niveles superiores de una agencia o un subcomponente organizativo. Una agencia también puede designar las cuentas de correo electrónico de otros empleados como FINALES cuando ocupan puestos en los que es probable que creen o reciban documentos de correo electrónico de conservación permanente (National Archives, 2013).

La mayoría de las agencias federales de EE.UU. cumplieron con el plazo de 2016, pero ‘manifestaron a NARA que no está claro cómo medirán su éxito y

sabrán que cumplen con los requisitos federales de gestión de documentos’ (Ogrysko, 2016). Los autores no creen que el enfoque de capturar solo el correo electrónico de los escalones más altos de una organización sea suficiente como método de gestión documental o archivístico. Sin embargo, pensamos que el reconocimiento implícito de que el valor archivístico reside en todas las colecciones de correos electrónicos, más que en una selección arbitraria de los usuarios, es un paso muy significativo.

Por lo tanto, el desarrollo de correos electrónicos como la principal herramienta de comunicación empresarial debería llevarnos a considerar si no estamos malinterpretando la naturaleza de los documentos digitales. Gran parte de la literatura (y hay mucha) parece basarse en la asunción de que los correos electrónicos “importantes” pueden y deben tener las mismas características que las cartas en papel. En otras palabras, son supuestamente autónomos, su contexto es simplemente evidente a partir de su contenido (o si no requiere a un humano que lo proporcione en forma de metadatos semánticos adicionales), y se pueden “archivar” cuidadosamente con documentos en una estructura precoordiada que tiene algún tipo de significado principal para la organización. Por tanto, se puede acceder a ellos a través de catálogos convencionales. Sin embargo, la evidencia que hemos citado sugiere que los datos electrónicos no son así en absoluto y no hay perspectivas de que lo sean.

Aunque hemos elegido el correo electrónico para ilustrar nuestro argumentación, no es un caso único. La llegada de Google Docs, SharePoint, las intranets y los tuits retuiteados ha creado una red de datos digitales adicionales (disfrazados como documentos) en el corazón de la mayoría de las modernas organizaciones. Hoy en día, los datos no son homogéneos, sino heterogéneos con una diversidad de objetos y enlaces, todos los cuales juntos constituyen un documento combinado que merece ser conservado. Durante 30 años o más, las organizaciones (y también los individuos) se han vuelto cada vez más codependientes de estos datos. Estos tipos de datos individuales (correos electrónicos, tuits, Google Docs) y los retos que suponen para las organizaciones y los archivos, son meramente un síntoma de la condición más amplia de la ‘datificación’.

## Conclusión

Como en todas las situaciones médicas con las que estamos familiarizados, los tratamientos que solo tratan los síntomas nunca proporcionarán una cura. Podemos desarrollar la ‘píldora para el correo electrónico’ y una ‘píldora para SharePoint’, pero el daño subyacente al cuerpo del archivo continuará a menos que cambie nuestro enfoque. Los archiveros están en una posición privilegiada para observar el progreso de la enfermedad y desarrollar una respuesta sistémica. Los

archiveros y sus usuarios, particularmente la comunidad académica, deben pasar a una situación en la que vean las colecciones archivísticas como colecciones de datos en línea que tienen características totalmente diferentes de las colecciones analógicas tradicionales. Su interpretación debe ser fluida y susceptible de ser analizada a través de toda una gama de nuevas y crecientes herramientas sofisticadas. Las viejas recetas de los metadatos, del orden original e incluso de los documentos originales han perdido su poder.

## Bibliografía

- Ahlberg, C., y Shneiderman, B. (1994). 'Visual information seeking: tight coupling of dynamic query filters with starfield displays,' en *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, eds B. Adelson, R. Dumais, y J. Olson (Boston, MA), 313-317.
- Allan, A. (2014). *Records Review*. London: Cabinet Office and The National Archives. Disponible en línea en: <https://www.gov.uk/government/publications/records-review-by-sir-alex-allan>
- Allan, A. (2015). *Government Digital Records and Archives Review*. London: Cabinet Office and The National Archives. Disponible en línea: <https://www.gov.uk/government/publications/government-digital-records-and-archives-review-by-sir-alex-allan>
- Allison, A., Currall, J., Moss, M., y Stuart, S. (2010). "Digital identity matters". *JASIST*, 56, 364-372. doi: 10.1002/asi.20112
- Brooks, C., y Warren, R. P. (1938). *Understanding Poetry: An Anthology for College Students*. New York, NY: Henry Holt and Company.
- Caswell, M. L. (2016). The Archive' is not an archives: on acknowledging the intellectual contributions of archival studies. *Reconstruction: Studies in Contemporary Culture* 16.
- Chowcat, I. (2015). *Spotlight on the Digital - Spotlight on the Digital Recent trends and Research in Scholarly Discovery Behaviour*. Bristol: JISC. Disponible en línea en: [https://digitisation.jiscinvolve.org/wp/files/2015/10/spotlight\\_literature\\_review\\_sept2015.pdf](https://digitisation.jiscinvolve.org/wp/files/2015/10/spotlight_literature_review_sept2015.pdf)
- Cohen, D. (2011). *Defining Digital Humanities*. *Research Without Borders Conference, Columbia University*. Disponible en línea en: <http://scholcomm.columbia.edu/2011/02/10/defining-the-digital-humanities/>
- Davis, G. (2011). *The Well-Wrought Textbook, A look back at Brooks and Warren's college classic, Understanding Poetry*, *Humanities* 32. Disponible en línea en: <https://www.neh.gov/humanities/2011/julyaugust/feature/the-well-wrought-textbook>
- Dearstyne, y Bruce, W. (1987). What is the use of archives? a challenge for the profession. *Am. Arch.* 50, 78-87. doi: 10.17723/aarc.50.1.572q383767657258
- Duff, W., y van Ballegoie, M. (2006). *Archival Metadata*. Toronto, On: University of Toronto. Disponible en línea en: <http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/archival-metadata>
- Email Management, (2016). National Archives. Disponible en línea en: <https://www.archives.gov/records-mgmt/email-mgmt>
- Evans, R. (2015). *Archives of War: Media, Memory and History Conference*. TNA, London. 30 November.
- Graham, S., Milligan, I., y Weingart, S. (2015). *Exploring Big Historical Data: The Historian's Macroscope*. London: Imperial College Press.
- Greenhalgh, M. (2014). *The Data Science in Government Programme: Progress So Far, Examples and Findings*. *Policy Lab*. Disponible en línea en: <https://openpolicy.blog.gov.uk/2014/09/18/data-science-2/>
- Grishman, R., y Sundheim, B. (1996). "Message understanding conference - 6: A Brief History," en *COLING '96 Proceedings of the 16th International Conference on Computational Linguistics*, Vol. 1, 466-471. Disponible en línea en: <http://www.aclweb.org/anthology/C96-1079>
- Guldi, J., y Armitage, D. (2014). *History Manifesto*. Cambridge: Cambridge University Press.
- History Hub (2015). NARA Disponible en línea en: <https://historyhub.history.gov/docs/DOC-1012>
- Hitchcock, T. (2014). *Big Data, Small Data and Meaning*. Historyonics. Disponible en línea en: <http://historyonics.blogspot.com/search?updated-max=2015-05-29T02:11:00-07:00&max-results=7&start=5&by-date=false>
- Hitchcock, T. (2015). *The UK Web Archive, Born Digital Sources and Rethinking the Future of Research*. Historyonics. Disponible en línea en: <http://historyonics.blogspot.com/2015/>
- Internet World Stats (2017). Disponible en línea en: <https://www.internetworldstats.com/emarketing.htm>
- Klimt, B., y Yang, Y. (2004a). "Introducing the Enron Corpus," en *CEAS 2004 - First Conference on Email and Anti-Spam (Mountainview, CA)*. Disponible en línea en: <http://nl.ijs.si/janes/wp-content/uploads/2014/09/klimtyang04a.pdf>
- Klimt, B., y Yang, Y. (2004b). "The enron corpus a new dataset for email classification research," en *Machine Learning: ECML 2004*, eds J. F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi (Berlin: Springer), 217-224.
- Lavoie, B. (2014). *The Open Archival Information System (OAIS) Reference Model: DPC Technology Watch Report 14-02 October 2014*, 2nd Edn. York: Digital Preservation Coalition. Disponible en línea en: <http://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file>
- Luca, M., Wu, T., Couvidat, S., Frank, D., y Seltzer, W. (2015). "Does google content degrade google search? Experimental evidence," in *Harvard Business School Working Paper*, 16-035 (Cambridge, MA: Harvard Business School). Disponible en línea en: [https://www.hbs.edu/faculty/Publication%20Files/16-035\\_2260fc69-1f63-466f-b4df-7957e77e2a3f.pdf](https://www.hbs.edu/faculty/Publication%20Files/16-035_2260fc69-1f63-466f-b4df-7957e77e2a3f.pdf)
- Lynch, C. A. (2003). "Colliding with the real World: heresies and unexplored questions about audience, economics, and control of digital libraries," en *Digital Library Use: Social Practice in Design and Evaluation*, eds A. Bishop, B. Butterfield, y N. Van House (Cambridge, MA: MIT Press), 191-216.
- Mayer-Schönberger, V. (2011). *Delete: The Virtue of Forgetting in the Digital Age*. Princeton, NJ: Princeton University Press.
- Merrin, W. (2014). *Media Studies 2.0*. London: Routledge.

- Moretti, F. (2000). Conjectures on World literature. *New Left Rev.* 1, 54-68. Moretti, F. (2013). *Distant Reading*. London: Verso.
- Moss, M. (2005). The hutton inquiry, the president of Nigeria and what the butler hoped to see. *Eng. Hist. Rev.* 120, 577-592. doi: 10.1093/ehr/cei121
- National Archives (2013). Bulletin 2013-02: Guidance on a New Approach to Managing Email Records. *NARA Bulletin*. Disponible en línea en: <https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html>
- Nicholas, D. (2007). "If we do not understand our users, we will certainly fail," en *The E-Resources Management Handbook*, ed R. Anderson (UKSG). Disponible en línea en: <https://www.uksg.org/publications/ermh>
- Ogrysko, N. (2016). *Most Agencies Say They'll Meet Year-End Records Management Deadline*. Federal News Radio. Disponible en línea en: <https://federalnewsradio.com/agency-oversight/2016/03/agencies-say-theyll-meet-year-end-records-management-deadline/>
- Padia, K. (2012). *Visualizing Digital Collections at Archive-It*. Dissertation/master's thesis, Old Dominion University, Norfolk, VA. Disponible en línea en: [https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1008&context=computerscience\\_etds](https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1008&context=computerscience_etds)
- Pitti, D. V. (2006). Technology and the transformation of archival description. *J. Arch. Organ.* 3, 9-22.
- Prabhakaran, V., Reid, E. E., y Rambow, O. (2014). "Gender and power: how gender and gender environment affect manifestations of power," en *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1965-1976*. Disponible en línea en: [https://cs.stanford.edu/~vinod/papers/EMNLP\\_genderpaper\\_final.pdf](https://cs.stanford.edu/~vinod/papers/EMNLP_genderpaper_final.pdf)
- Rea, S. (2015). *Six Degrees of Francis Bacon*. Carnegie Mellon University. Disponible en línea en: <http://www.cmu.edu/news/stories/archives/2015/october/francis-bacon-launch.html>
- Rosenthal, D. (2017). *DSHR's Blog*. Disponible en línea en: <http://blog.dshr.org/search/label/storage%20costs>
- Schofield, T., Kirk, D., Amaral, T., Schofield, G., y Ploetz, T. (2015). Archival Liveliness: Designing With Collections Before and During Cataloguing and Digitization. *Digital Humanities Quarterly*. 9:3. Disponible en línea en: <http://www.digitalhumanities.org/dhq/vol/9/3/000227/000227.html>
- Schott Syme, H. (2003). Becoming speech: voicing the text in early modern English courtrooms and theatres. *Compar(a)ison*, 11, 107-124.
- Shabou, B. M., y Sokhn, M. (2017). "The new information technologies at the service of historical and cultural heritage and tourism promotion," en *Integrating ICT in Society*, eds Iana Atanassova et al., (Zagreb: FF Press), 219-234.
- Smith, C. (2017). *The Pixel 2' Unlimited Photo Storage isn't Exactly Unlimited*. BGR. Disponible en línea en: <http://bgr.com/2017/10/06/pixel-2-features-unlimited-photo-storage/>
- TNA (2016). *The Application of Technology-Assisted Review to Born Digital Records, Transfer, Inquiries and Beyond- Research Report*. London: The National Archives. Disponible en línea en: <http://discovery.nationalarchives.gov.uk/results/r?q=Records+of+the+UK+Atomic+Energy+Authority%27s+production+divisions+touching+all+aspects+of+atomic+energy+research%2C+day+to+day+procedures%2C+administrative+functions+and+industrial+relations>
- TNA (2018), *Online Catalogue*. Disponible en línea en: <http://discovery.nationalarchives.gov.uk/results/r?q=Records+of+the+UK+Atomic+Energy+Authority%27s+production+divisions+touching+all+aspects+of+atomic+energy+research%2C+day+to+day+procedures%2C+administrative+functions+and+industrial+relations>
- Turkel, W. J., Kee, K., y Roberts, S. (2013). "A method for navigating the infinite archive," in *History in the Digital Age*, eds Toni Weller (London, Routledge), 61-75.
- Turnbaugh, R. C. (1983). Living With a Guide. *Am. Arch.* 46, 449-452.
- Underwood, T. (2014). Theorizing research practices we forgot to theorize twenty years ago. *Representations* 127, 64-72. doi: 10.1525/rep.2014.127.1.64
- Viégas, F. B., Golder, S., y Donath, J. (2006). "Visualizing email content: portraying relationships from conversational histories," en *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY: ACM).
- Walter, C. (2005). Kryder's Law-The doubling of processor speed every 18 months is a snail's pace compared with rising hard-disk capacity, and Mark Kryder plans to squeeze in even more bits. *Sci. Am.* 293, 32-33. doi: 10.1038/scientificamerican0805-32
- Waugh, A. (2014). *Email - a Bell Weather Records System*. Recordkeeping Roundtable. Disponible en línea en: <http://rkroundtable.org/2014/06/30/email-a-bellwether-records-system/>
- Weinberger, D. (2011). *Too Big to Know*. New York, NY: Basic Books.
- Zachte, E. (2011). *Wikipedia Visualisations*. Erik Zachte's Wikipedia / Wikimedia Portfolio. Disponible en línea en: <http://infodisiac.com/Wikimedia/Visualizations/>

## Notas

<sup>1</sup> Una versión abreviada de este artículo se presentó en la conferencia Activation and impact: the societal role of records and record-keepers (FARMER) Conference 2016 del Forum on Archives and Records Management Education and Research en la Dundee University en abril de 2016. Nos gustaría agradecer a los organizadores el habernos invitado a hablar y a Willaim Vinh-doyle y Daniel German por sus comentarios.

<sup>2</sup> NT. La zona de Whitehall es el centro administrativo del Gobierno británico,

<sup>3</sup> Nuix, <http://www.nuix.com/>

<sup>4</sup> Archive-It, <https://archive-it.org>

<sup>5</sup> California Digital Library, <http://www.cdlib.org>

<sup>6</sup> Library of Congress Archived Web Sites, <http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>

<sup>7</sup> Pandora, <http://pandora.nla.gov.au>

<sup>8</sup> Themail, <http://alumni.media.mit.edu/~fviegas/projects/themail/study/index.htm>

<sup>9</sup> Virtual St. Paul's Cathedral Project, <https://vpcp.chass.ncsu.edu/>

<sup>10</sup> Old Bailey Voices, <https://oldbaileyvoices.org/>

<sup>11</sup> (The) Digital Panopticon-The Global Impact of London Punishments. 1780-1925, Disponible en línea en: <http://www.digitalpanopticon.org/>