



GEOFFREY YEO

¿Podemos mantenerlo todo?

El futuro de la valoración en un mundo de profusión digital

Can we keep everything?

The future of appraisal in a world of digital profusion

Geoffrey Yeo

g.yeo@ucl.ac.uk

Honorary researcher in archives and records management at University College London (UCL)

Citación: Yeo, Geoffrey (2020). "¿Podemos mantenerlo todo? El futuro de la valoración en un mundo de profusión digital". *Tábula*, n. 23, pp. 153-171

Traducción: Verónica Fernández y Luis Hernández Olivera

Versión original: "Can we keep everything? The future of appraisal in a world of digital profusion", Brown, Caroline (ed) *Archival Futures*, London, Facet Publishing, 2018.

Traducido y publicado con la autorización del autor y Facet Publishing. Contact pete.baker@facetpublishing.co.uk for rights enquiries

Recibido: 20-7-2020. *Aceptado:* 28-10-2020

DOI: <https://doi.org/10.51598/tab.845>

Resumen analítico / Analytic summary

En este artículo nos preguntamos si podríamos comenzar a desarrollar una visión de los archivos digitales como un lugar donde nunca se destruye nada. Hoy en día, los defensores del 'big data' mantienen que la valoración y la selección se están quedando obsoletas. ¿Podemos –o debemos– intentar quedárnoslo todo como sugieren ahora, muchos expertos informáticos? ¿Qué lugar ocuparán las prácticas de selección y destrucción en un futuro mundo digital donde almacenar todo puede parecer más sencillo que acarrear con los gastos y complejidades que implica la valoración?

VALORACIÓN | BIG DATA | PRESERVACIÓN DIGITAL | RETENCIÓN | CAMBIO TECNOLÓGICO

This text asks whether we might begin to develop a vision of digital archives where nothing is ever destroyed. Today, proponents of 'big data' often argue that appraisal and selection are obsolescent. Can we – should we – try to keep everything, as many computing specialists now suggest? What place will selection and destruction practices have in a future digital world where storing everything may seem easier than incurring the costs and complexities of appraisal?

APPRAISAL | BIG DATA | DIGITAL PRESERVATION | RETENTION | TECHNOLOGICAL CHANGE

Retención selectiva

Desde los primeros días de la creación de documentos se han tomado decisiones relativas a la destrucción. En otras épocas se llegó a pensar que los documentos estaban destinados a ser permanentes y que los custodios de dichos documentos no podían participar en su destrucción, hasta que la práctica de conservar todo tuvo que ser abandonada debido a las ingentes cantidades físicas de documentos acumulados durante el siglo XX. Las prácticas de destrucción sistemática, sin embargo, tienen una historia mucho más larga. En los palacios y las ciudades del mundo antiguo, los documentos contables se solían destruir de forma rutinaria al año de su creación (Eidem, 2011, 13; Palaima, 2003). En 1731, en los Archivos Reales de Cerdeña se dio la orden de destruir todos 'los papeles inútiles' (Ducheyn, 1992, 17); tres décadas más tarde, el archivero francés Pierre-Camille Le Moine insistía de manera similar en la destrucción de ítems aparentemente triviales (Delsalle, 1998, 154). En el Reino Unido, la Public Record Office Act de 1877 autorizó la eliminación de documentos que no fueran lo suficientemente valiosos (Cantwell, 1991, 277-80). Las metodologías de la valoración del siglo veinte articuladas por Theodore Schellenberg (1956) formalizaron y alentaron estas prácticas, pero los orígenes de dicha selección son mucho más antiguos.

Antes de que la tecnología digital deviniera disponible de manera generalizada, los archiveros solían presentar dos argumentos a favor de la destrucción selectiva: la necesidad de ahorrar el costoso espacio de almacenamiento y la dificultad de encontrar y utilizar material 'importante' si los archivos estaban congestionados con documentación efímera presuntamente sin valor. Estos argumentos aparecen de forma muy similar en la obra de Charles Johnson a comienzos del siglo XX y en los escritos de Gerald Ham en la última década del siglo (Ham, 1993, 3; Johnson, 1919, 43-4). Aunque Hilary Jenkinson (1922) se

negó a permitir que el archivero participara de alguna forma en la selección, muchos autores posteriores vieron un papel específico para los archiveros a la hora de ‘dar forma a un documento histórico gestionable’ (Ham, 1993, 3). Los primeros gestores de documentos afirmaban que ‘el propósito de la gestión de documentos puede establecerse en términos simples: *menos y mejores documentos*’ (Leahy and Cameron, 1965, 17). Además los archiveros solían pensar que una menor cantidad de documentos equivalía a mejores documentos. Para algunos, el volumen del material presentaba problemas que solo podían superarse con una ‘selección rigurosa’ (Hall, 1908, 3); para otros, la tarea del archivero era ‘conservar la imagen más nítida posible de la sociedad contemporánea... eligiendo los mejores documentos’ (Cook, 1991, 33).

A principios del siglo XXI, estos argumentos fueron trasladados a veces al nuevo mundo de los recursos digitales. Adrian Cunningham (2008, 535) se opuso a la conservación de ‘montañas de datos anárquicos y sin gestionar’, a favor del ‘compromiso de documentar las cosas importantes que suceden en la sociedad y en la administración pública’. Ross Harvey (2008) también abogó a favor del método selectivo, basándose en que ‘la selección y la valoración son claves para asegurar que los datos y los documentos científicos sean utilizables y reutilizables’. Según Harvey, la selección es una respuesta necesaria a los costes de almacenamiento y mantenimiento de grandes cantidades de datos digitales y ofrece beneficios con respecto a la calidad de la conservación y a la capacidad de los archiveros para demostrar la confiabilidad de sus fondos.

Sin embargo, han existido voces disidentes, especialmente entre aquellos que han examinado las prácticas de valoración y selección documental desde la perspectiva de otras disciplinas. Entre los críticos más contundentes se encuentra el experto Thomas Tanselle, quien escribió que los archiveros:

... con frecuencia se arrojan el deber, para con la sociedad y el futuro, de eliminar el material insignificante. La arrogancia de esta postura es asombrosa. No hay manera de saber qué instrumentos otra persona, ahora o en el futuro, considerará importantes; y existen cientos de ejemplos de materiales que fueron ignorados en un momento y considerados de gran valor en otros. (Tanselle, 2002, 258).

Aunque las críticas de Tanselle tuvieron una acogida hostil (Cox, 2004, 267-8), los archiveros eran conscientes de que tanto los objetivos como los métodos de valoración son muy controvertidos. Muchos archiveros han subrayado que los criterios de valoración y selección son arbitrarios y que los ejercicios de evaluación pueden ser impredecibles, inconsistentes y estar llenos de prejuicios. Llevar a cabo una valoración evaluando funciones –propuesto por varios de los pensadores líderes desde la década de los 90 (Cook, 1992; Robyns, 2014)– no parece menos subjetivo que evaluar directamente los documentos de archivo.

Para Mark Greene (2002), la valoración es un arte, no una ciencia. Para otros la convicción de que el objetivo de la valoración es el de identificar los documentos que tratan sobre ‘las cosas importantes que suceden en la sociedad» (Cunningham, 2008, 535) o ‘la deseada imagen total de la sociedad que debería dejarse para la posteridad’ (Cook, 1991, 33) está plagada de dificultades; muestra a los archiveros luchando por el poder de decidir lo que es ‘importante’ y así manipular la futura percepción del mundo.

Una nueva perspectiva

Para los tecnólogos, un problema crítico es que el almacenamiento digital se ha convertido en algo, si cabe, todavía más económico. El coste de almacenamiento en un disco local en 2010 supuso apenas un 1% de lo que era en el 2000 y el 0,001% de lo que había sido en 1990 (Komorowski, 2014). Hoy en día, los servicios en la nube nos ofrecen almacenamiento digital a un coste todavía inferior o incluso aparentemente gratis. Ya en 2004, Google ofreció a los usuarios de su servicio Gmail grandes cantidades de almacenamiento ‘gratis’ diciéndoles que ya no necesitarían borrar ningún correo electrónico (Kirschenbaum, 2008, 98). En 2015, anunció que ofrecería almacenamiento ilimitado a la mayoría de los usuarios de Google Apps (más tarde denominado G-Suite). Cuando los costes de almacenamiento a gran escala son insignificantes –o al menos tan bajos que cualquier servicio en la nube puede incluirlos disimuladamente en el precio cobrado por otros servicios– llegamos a un punto en el que, en la mayoría de los casos, los costes iniciales de almacenamiento no tienen ninguna influencia en las decisiones tomadas sobre la conservación.

Uno de los primeros campos de investigación en informática que sirvió de base para el concepto de un almacenamiento digital cada vez más barato y más amplio fue el ‘*lifelogging*’ o el registro de las actividades vitales: el uso de herramientas tecnológicas para capturar y conservar documentos supuestamente completos de las vidas individuales. Originalmente basado en sensores y cámaras de bolsillo, las iniciativas *lifelogging* se difundieron imaginando un futuro en el que la totalidad de las publicaciones en las redes sociales, los mensajes de texto, los correos electrónicos y las conversaciones telefónicas de un individuo se capturaban y almacenaban junto con datos de ubicación, imágenes, sonidos y mediciones ambientales. Estos registros vitales podrían mejorarse aún más mediante la integración del llamado ‘internet de las cosas’, ofreciendo el potencial de registrar todas las interacciones que experimentamos. Sus defensores sostuvieron que todos estos documentos se podían conservar porque los costes de almacenamiento serían mínimos. Las primeras iniciativas de *lifelogging* tuvieron poca aceptación, pero los investigadores informáticos han seguido trabajando en este campo y en la

actualidad suelen recomendar *lifelogs* para el bienestar personal, la terapia de reminiscencia o el tratamiento de la amnesia (Isaacs et al., 2013; Jones, 2014, 7-10).

Por otro lado, lo que sí ha generado una mayor expectación son los datos masivos o *'big data'*: el crecimiento exponencial de los volúmenes de datos digitales producidos y almacenados como consecuencia de la proliferación de dispositivos informáticos en los gobiernos, las industrias y el público en general. El concepto de big data tiene una gran repercusión, entre otras cosas por su potencial de investigación de mercado, toma de decisiones y ventajas competitivas. Se caracteriza no solo por la abundancia de datos, sino también por el uso de técnicas informáticas innovadoras para analizar datos a una escala antes impracticable, con la expectativa de descubrir patrones, tendencias o relaciones que de otro modo ignoraríamos (Ernst and Young, 2014; Information Resources Management Association, 2016; Müller et al., 2016). Además de su uso en el mundo empresarial, estas técnicas analíticas ofrecen nuevas oportunidades para la investigación académica (Thomas y Johnson, 2015).

En el siglo XXI, el uso del análisis de datos ha dado al traste con el pensamiento tradicional sobre la necesidad de selección; en lugar de buscar reducir los datos a cantidades manejables, los analistas afirman ahora que mayores volúmenes de datos producen mejores resultados y que estas nuevas técnicas serán más eficaces si se conserva todo y está disponible al análisis. Como explicaron Kenneth Cukier y Viktor Mayer-Schoenberger, el mundo del big data nos exige:

Recoger y utilizar una gran cantidad de datos en lugar de conformarnos con pequeñas cantidades...Durante la mayor parte de la historia, se ha trabajado con cantidades relativamente pequeñas de datos porque las herramientas para recopilar, organizar, almacenar y analizar la información eran pobres. Se cribaba la información fiable al mínimo para poder examinarla con mayor facilidad... hoy... no dependemos de las muestras pequeñas. (Cukier y Mayer-Schoenberger, 2013)

En el pasado se nos decía que las habilidades cognitivas humanas fijaban los límites de la cantidad de datos que se podían asimilar, pero ahora una conservación a gran escala resulta apropiada en un mundo donde los programas informáticos poseen una mayor capacidad de procesamiento. Los defensores de este método insisten en que 'nunca se deben desechar datos' y hablan de organizaciones que emplean *'data lakes'* 'lagos de datos digitales' infinitamente expandibles que les permitirán aprovechar todos sus datos utilizando la búsqueda, la minería de datos, la minería de texto y otras herramientas analíticas derivadas de, o mejoradas por la inteligencia artificial (Davenport y Kim, 2013, 8; Losey, 2015).

A menudo los profesionales de los archivos han considerado estas ideas poco satisfactorias. Según Terry Cook (1991, 33), 'incluso si los archiveros pudieran quedárselo todo, no deberían hacerlo'. Richard Cox sostuvo que la evaluación

sigue siendo necesaria en entornos digitales y calificó de 'fantasía' la idea de que la selección es innecesaria (Cox, 2011; Cox y Larsen, 2008, 324). Reiterando las preocupaciones eternas sobre el volumen y manejabilidad, Lawrence Serewicz (2010), Janet Knight y Kate Cumming (2012) también se opusieron a las propuestas de conservar todo, afirmando que los volúmenes de datos sobrepasarían lo manejable. Kevin Dale (2015) sugirió que los gestores de documentos podrían echar una mano a los analistas de 'big data' eliminando datos 'extraños'. Los profesionales de los documentos, como observó el historiador francés Pierre Nora (1989, 14), han 'aprendido que la esencia de su profesión es el arte de la destrucción controlada', y son comprensiblemente reacios a abandonar dicho arte que a menudo parecía fundamental en la definición de su identidad profesional.

Algunas voces dentro de la profesión archivística, sin embargo, han sugerido que las creencias tradicionales sobre la necesidad de valoración y destrucción deben volver a revisarse y que no podemos seguir considerando la idea de mantener todo como un capricho o algo inasequible. El gestor de documentos Steve Bailey (2008, 100-2) comentó que el énfasis sobre la retención selectiva resulta ajeno tanto a la industria informática como a la cultura popular, y que negarse a conservarlo todo implica nadar contra poderosas corrientes de la sociedad. Anne Gilliland (2014, 54) siguió los argumentos de Bailey y señaló que mantener todo ofrecería un mejor soporte a los nuevos modos de investigación. En un artículo titulado '¿Por qué nos dirigimos a un mundo donde nos vamos a quedar con todo para siempre?', Barclay Blair (2015) afirmó que una conservación total está en marcha, independientemente de si los archiveros lo aprueban o no.

Los archiveros y gestores documentales no pueden ignorar las tendencias tecnológicas y sociales que identificaron Bailey y Blair. Esta petición de mantener todo no va a desaparecer sin más; de hecho, probablemente dicha exigencia se vuelva más fuerte. En los próximos años asistiremos con seguridad a más debates sobre la retención, en los que los profesionales de los archivos figurarán inevitablemente. Si bien es posible que todavía exista alguna preocupación por la capacidad para gestionar enormes cantidades de documentos, es el momento de considerar la posibilidad de que la valoración, como la conocíamos en el siglo XX, se está acercando a su fin.

Viabilidad del cambio

La expectativa de no destruir nada en los archivos podría suponer un desafío e incluso podría parecer aterrador. Conceptualmente, sin embargo, hay muchas razones para elogiarla. Las agrupaciones archivísticas intactas –una agrupación que suponga un verdadero fondo de archivo– no supone la pérdida del contexto documental interno (Yeo, 2012, 79). Si es cierto que cada uno de los documentos

de una agrupación documental tiene redes de relaciones con otros documentos y que 'la eliminación de cualquiera de ellos perjudica la integridad de los archivos' (Duranti, 2015, 28), conservar todo impediría dicho daño. Para todos aquellos preocupados por la falibilidad de la valoración, serviría para eliminar la necesidad de tomar decisiones subjetivas. Para los que buscan la integración archivística, pondría fin al privilegio de ciertas voces –y el silenciamiento de otras– que ocurre cuando los archiveros deciden preservar los documentos de forma selectiva.

Puede que haya también razones más pragmáticas para conservarlo todo. Como observaron los pioneros de la gestión de documentos, 'la selectividad es una larga y ardua tarea' (Leahy y Cameron, 1965, 228) por lo que parecería adecuado plantearse si en la actualidad preservar todo en forma digital podría resultar menos costoso y menos intensivo a la hora de emplear recursos que la tarea de una detallada valoración. Por supuesto, el mayor gasto de conservación no es el coste de los soportes de almacenamiento, sino el coste de programas de conservación activos y continuos; sin embargo, algunos expertos en preservación sostienen que las barreras de coste de estos programas son inferiores a lo que se solía pensar (Gollins, 2009; Rosenthal, 2010). Aunque en la década de los 90 se creía que la preservación digital costaría más que conservar papel, al menos una iniciativa de preservación durante la década de 2010 ha descubierto que los costes unitarios de la conservación digital son inferiores a los del papel (Mumma, Dingwall y Bigelow, 2011, 116).

También se ha descubierto que eliminar los rastros digitales no es fácil. Copias idénticas o casi idénticas de documentos digitales proliferan a menudo en diferentes soportes de almacenamiento, y cuando se destruye una copia rara vez es posible asegurar que no existen otras. Además las técnicas forenses pueden resucitar documentos que se creían eliminados. Los escépticos podrían preguntar por qué seguimos incidiendo en los costes de valoración a la hora de destruir documentos, cuando los entornos tecnológicos hacen imposible, a menudo, asegurar su eliminación completa.

Sin embargo, un mundo en el que se guarda todo sería un mundo con masivas cantidades de contenido digital. Se dice que los volúmenes de datos crecen a un ritmo del 40% anual (Information Resources Management Asociación, 2016, 1862) y que, para 2020, cada persona en el mundo generará alrededor de 1,7 megabytes cada segundo (Kumar, 2017). Estas cifras son aproximativas, pero las tendencias indican que son certeras. Necesariamente, los archiveros se cuestionarán si será factible abandonar las prácticas de selección cuando las futuras cantidades de material excedan todo lo que hemos experimentado en el pasado.

Y aun siendo los costes unitarios de preservación bajos, es improbable que seamos capaces de preservar grandes cantidades de documentos digitales a menos que se introduzcan técnicas de ahorro de trabajo. Aumentar las prácticas que dependen en gran medida del esfuerzo manual no es una opción probable. Sin

embargo, los expertos en conservación digital reconocen la necesidad de minimizar la dependencia de la intervención humana y están trabajando para automatizar las tareas de conservación. Se han desarrollado herramientas para automatizar tareas tales como la identificación del formato de archivo, la migración, la corrección de errores y la comprobación de la estabilidad, que cada vez más se agrupan en paquetes integrados para simplificar los flujos de trabajo de conservación. Y como los especialistas en preservación son conscientes de que estas herramientas y flujos de trabajo pueden descomponerse cuando se enfrenten a colecciones grandes o complejas (King et al., 2012), están explorando nuevos métodos automatizados que proporcionen una mayor adaptabilidad, lo que incluye el uso de la potencia de la computación de alto rendimiento (Arora, Esteva y Trelogan, 2014; Open Preservation Foundation, 2016). Los investigadores están también indagando en la viabilidad de automatizar los procesos de detección, planificación y gestión de riesgos que exige la preservación (Becker, Faria y Duretec, 2014; Waddington et al., 2016).

La mejor forma de describir los métodos de conservación actuales es como semiautomatizados, pero en el futuro podemos esperar ver tecnologías más maduras que permitan la preservación a una escala mucho mayor. Aunque en ocasiones se ha considerado la computación de alto rendimiento y la conservación digital a gran escala como ambientalmente insostenibles, los informáticos están desarrollando de forma activa nuevas formas de reducir el impacto medioambiental de la tecnología (Rahman, 2016). Algunos expertos en preservación digital prevén un tiempo aún lejano cuando los objetos digitales autogestionarán en gran medida su propia preservación y apenas se necesitará la aportación humana, excepto para el establecimiento de políticas y la supervisión.

Como a menudo han señalado los defensores de la selección, no sirve de nada conservar grandes cantidades de documentos si son difíciles de encontrar o son incomprensibles; el suministro de herramientas de recuperación, interpretación y contextualización de los documentos digitales del futuro supondrá todavía un desafío mayor. Los archiveros tienen ahora herramientas automatizadas o semiautomatizadas capaces de extraer los metadatos técnicos necesarios para la preservación y accesibilidad digital, pero para que la retención a gran escala sea factible, también necesitarán tecnología capaz de generar metadatos descriptivos o encontrar otra forma de proporcionar una posibilidad de descripción tradicional. El trabajo descriptivo manual no tendrá la capacidad de mantener el ritmo de los grandes volúmenes de documentos digitales y harán falta nuevas estrategias informáticas.

Parece probable pues que, para conseguir gran parte de esta capacidad, los archiveros buscarán implementar las herramientas analíticas de big data mencionadas anteriormente. Los servicios de almacenamiento en la nube ya han comenzado a ofrecer a sus clientes herramientas analíticas mejoradas por la inteligencia

artificial (o 'computación cognitiva') y técnicas de aprendizaje automático, y muchos comentaristas creen que estas herramientas se están convirtiendo en suficientemente poderosas como para servir de base para unos archivos digitales 'inteligentes' y una gestión documental expansible (Shinkle, 2017). A un nivel básico, pueden buscar texto, identificar términos temáticos, nombres y fechas, mapear su frecuencia y descubrir patrones y relaciones entre ellos; otras herramientas más avanzadas pueden intentar detectar referencias a actividades o acontecimientos e identificar temas no explícitos en el contenido textual. Algunas tienen como objetivo reconocer el contenido de las grabaciones de voz o audiovisuales, u objetos en imágenes fotográficas. Muchas ofrecen la posibilidad de utilizar los resultados del análisis de contenido para clasificar de manera automática los documentos digitales según categorías jerárquicas como las utilizadas en los tradicionales cuadros de clasificación (Warland y Mokhtar, 2013), aunque se sabe cada vez más que imponer una única organización no funciona en los dominios digitales, donde los documentos se pueden agrupar y reagrupar con diferentes propósitos (Yeo, 2017, 180-2). Para ayudar a los usuarios a explorar las agrupaciones grandes y complejas, se puede generar una variedad de vistas a través de técnicas de visualización (Whitelaw, 2015). La documentación de las interrelaciones de los documentos, y de su contexto o procedencia, también puede ser semiautomatizadas utilizando nuevos métodos computacionales (Lemieux, 2016; Spencer, 2017; Yeo, 2013).

Aunque muchas de estas tecnologías en la actualidad están verdes, se están desarrollando con gran rapidez; varios estudios recientes han sugerido que la recuperación de datos asistida por ordenador puede ser más precisa que el método manual (Gricks y Ambrogio, 2015). Sin embargo, esta generación necesita una gran 'formación' en sistemas de aprendizaje automático para lograr resultados satisfactorios. Los sistemas más desarrollados están diseñados para asistir al descubrimiento electrónico y la investigación acerca de su reutilización en archivos todavía está en curso (The National Archives, 2016).

Es también evidente que estas tecnologías tendrán un impacto cultural importante, tanto para los archiveros como para los usuarios. En el caso de los archiveros, no solo la adquisición de las habilidades necesarias para aplicar estrategias computacionales al procesamiento de archivos supondrá un desafío, sino también la integración de ideas computacionales con el pensamiento archivístico (Marciano, 2016); muchos archiveros tendrán reticencias a la hora de confiar a los ordenadores tareas que han pertenecido durante mucho tiempo al campo del juicio humano. Para los usuarios de archivos, las estrategias computacionales ofrecen nuevas oportunidades de acceso e investigación, pero muchas de estas oportunidades probablemente atraerán más a los investigadores que trabajan en tradiciones cuantitativas que a aquellos con intereses cualitativos o interpretativos.

Por supuesto, no podemos prever totalmente cómo será el desarrollo de estas tecnologías o cómo los usuarios responderán a ellas. Muy probablemente, se requerirán interfaces más transparentes y fáciles de usar antes de que se puedan ganar la confianza y aceptación generalizadas. Sin embargo, podemos esperar que, a medida que las tecnologías maduren, superarán cada vez más las dificultades propias de gestionar y utilizar documentos digitales a gran escala. Si bien todavía se necesitarán aportaciones humanas, un mayor nivel de confianza en las herramientas informáticas nos ayudará a controlar la avalancha digital facilitando un trabajo que de otro modo sería tremendamente laborioso.

¿Conservación transformada?

Volvamos pues a las preguntas planteadas al comienzo de este trabajo. En un mundo de profusión digital, donde la retención de documentos a mucha mayor escala se está haciendo realidad y donde muchos comentaristas ven la idea de conservar todo como algo realmente deseable, ¿cuál será el futuro de la valoración y la selección? En los archivos del mañana, ¿aspiraremos realmente a conservarlo todo?

A primera vista, todavía parece improbable una conservación total. Casi inevitablemente, al igual que en el mundo pre-digital, se seguirán perdiendo documentos a causa de fallos o de la poca predisposición a conservarlos. Como Catherine Marshall (2008) observó, seguirá siendo natural 'perder cierta cantidad de objetos digitales a causa de una negligencia benigna'.

Al menos en el corto y medio plazo, las pérdidas también seguirán siendo el resultado de la toma de decisiones calculadas. Todavía cuesta imaginar el momento en el que podamos de verdad escapar de la necesidad de restringir la preservación debido a limitaciones en la potencia informática, la capacidad de almacenamiento o a los recursos de gestión. Aunque la retención no selectiva será cada vez más común, debido a factores económicos, ciertas decisiones de selección seguirán siendo inevitables. Basta solo con tener en cuenta algunos ejemplos de 'big' data del mundo actual:

- En 2012, el CERN (Organización Europea para la Investigación Nuclear) procesó cada segundo 300.000 megabytes de datos de su Gran Colisionador de Hadrones; en años posteriores estas cantidades han sido incluso superiores. El CERN aplica algoritmos de selección de datos en el momento de procesarlos; en 2012, solo se conservó el 0,1%. Sin embargo, la cantidad de datos conservados ascendió a 15 millones de gigabytes en tan solo un año (Borek et al., 2014, 34). Dado que mantener el otro 99,9% cada año habría supuesto especulativamente que aumentara su almacenamiento –en más de diez centros de datos de toda Europa– a un equivalente de más

de 10.000 centros de datos, debemos poner en duda que incluso los defensores más entusiastas de la preservación hubieran podido persuadir al CERN para que aceptara quedárselo todo.

- Se ha estimado que el 86% de los mensajes de correo electrónico son correo 'spam' no deseado y que al día se envían alrededor de 400 mil millones de correos basura (Robertson, 2016). Incluso si estos mensajes pudieran ser de utilidad para algún proyecto de investigación en algún momento del futuro, parecería difícil encontrar una justificación económica para conservarlos.
- Varios años después de aceptar 'preservar cada tweet publicado', la Biblioteca del Congreso de los Estados Unidos se dio cuenta de que seguía siendo incapaz de dar acceso público a su 'archivo de Twitter'. El número de tweets superó enormemente las expectativas de la biblioteca y los problemas técnicos a la hora de gestionarlos y de hacerlos accesibles superaron con creces sus recursos (McGill, 2016).

En un futuro inmediato, no conservaremos todo lo que se cree digitalmente. No consideraremos que valga la pena configurar cada base de datos de manera que cada cambio sea capturado para la posteridad, y la preservación a gran escala exigirá más recursos de la que la mayoría de las instituciones archivísticas probablemente posean. El volumen de material digital todavía requerirá algún tipo de actividad de valoración para determinar lo que se debe preservar, incluso si (como en el caso del CERN) son algoritmos en lugar de agentes humanos los que lleven a cabo dicha evaluación. Por supuesto, algunos archiveros podrían argumentar que los tweets, correos electrónicos no deseados y los datos experimentales del CERN no deberían considerarse documentos de archivo. Ciertamente no son el prototipo de lo que en otros lugares es considerado como tal, es decir, el tipo de documento que la mayoría de nosotros asociamos a nuestra imagen de fondo de archivo (Yeo, 2008). Pero bien es cierto que cuestionarse la naturaleza de un documento de archivo ya no es una pregunta con una respuesta clara y decidir si un tipo particular de objeto digital es o no un documento de archivo, ya es, de por sí, un tipo de decisión de valoración.

Sin embargo, los archiveros tendrán que tratar de conservar una mayor cantidad de documentos en el futuro de los que han sido capaces de preservar en el pasado. Con el tiempo, según vayan haciéndose más sofisticadas y más accesibles las herramientas tecnológicas que asistan la preservación, descripción y el acceso, casi con toda seguridad disminuirán los problemas del tipo de los que se enfrentó la Biblioteca del Congreso con el 'archivo de Twitter'. La intervención humana seguirá siendo necesaria, pero se centrará más en aquellos aspectos de la conservación que dependen de un apoyo personalizado o niveles más altos de la creación de sentido. Cuando esto suceda, la retención a gran escala

ya no resultará tan desalentadora —especialmente lo que respecta a documentos personales u organizativos prototípicos— y conservarlos todos, en lugar de conservar solo algunos, se convertirá en una práctica profesional normal si no universal.

Las prácticas de valoración no desaparecerán, pero sus objetivos y aspectos serán diferentes. En muchos entornos, la valoración estará más centrada en determinar qué documentos se deben crear (qué tipo de hechos deberíamos documentar; qué tipo de preguntas o compromisos deben ponerse por escrito) que en determinar qué documentos se deben seleccionar para su conservación. Los archiveros continuarán encargándose de verificar la creación y protección correcta de los documentos de archivo. Además tendrán que garantizar que la rendición de cuentas y la transparencia no se vean obstaculizadas por sujetos u organizaciones que evadan la creación de archivos o que oculten o destruyan documentos con el fin de evitar su responsabilidad o la divulgación de información incómoda. En la mayoría de las organizaciones, también seguirá teniendo espacio la tarea que podríamos denominar valoración para determinar qué documentos deben trasladarse desde los sistemas operativos al almacenamiento inactivo o para identificar qué documentos precisan seguridad adicional. Algunas de estas tareas pueden ser automatizables, pero siguen siendo tareas de valoración —o algo muy similar a la valoración— que los archiveros desearán supervisar.

En muchas instituciones archivísticas, podemos esperar que la valoración se ocupe de la identificación de sistemas de gestión de documentos —o fondos, si optamos por una terminología más tradicional— para su incorporación; también podemos esperar que se acepten muchos más ingresos que en el pasado. Puede que algunas instituciones dedicadas a la recopilación de documentos ya no necesiten rechazar ninguna colección digital que se les ofrezca y que de hecho puedan 'conservarlo todo'. Incluso donde la aceptación siga siendo selectiva, veremos valoración y selección, *de* fondos en lugar de *dentro* de los fondos. No habrá lugar a expurgar fondos o para la mayoría de otras formas de valoración a nivel de ítem o de unidad de instalación. El muestreo, a menudo defendido en el pasado como método de preservación selectiva (Cook, 1991; Neumayer y Rauber, 2007), también se convertirá en una rareza; probablemente todavía será necesario cuando se creen documentos o datos a escala industrial, como en el caso del CERN, pero el muestreo de documentos procedentes de procesos administrativos (conocido por los archiveros como 'expedientes' o 'documentos de instancias particulares') se quedará anticuado.

Una excepción importante a esa visualización de la conservación casi total tiene que ver con las cuestiones de privacidad y confidencialidad. Puede que la legislación en materia de privacidad —como las leyes europeas de protección de datos— obligue a la destrucción de documentos relativos a individuos identificables. Este movimiento hacia la conservación total debe ser reprimido por una

legislación de este tipo, así como por el creciente reconocimiento de que los individuos en las sociedades modernas pueden, lo que supone un problema mayor, tener ‘derecho al olvido’. Como observó Gilliland (2014, 54), seguirá habiendo ‘razones aceptables’ que justifiquen una destrucción selectiva, especialmente cuando cuestiones de derechos humanos estén en juego. Cuando se considere que faltan requisitos legales o morales a la hora de ejecutar la destrucción, puede que todavía siga haciendo falta procedimientos que nos ayuden a identificar aquellos documentos considerados sensibles, y en lugar de destruirlos, ocultarlos o bien ponerlos a buen recaudo durante el tiempo que se estime necesario. Las instituciones que se enfrentan a la tarea de descubrir contenido sensible dentro de fondos digitales a gran escala recurren cada vez más a las técnicas computacionales que les ayuden en la evaluación (Baron y Borden, 2016; Moss, 2015, 9-11).

Finalmente, si así lo desearan, los archiveros que siguen considerando necesaria la valoración en la identificación y selección de ‘las pocas joyas que tienen valor’ (Cook, 1991, 45) entre montañas de documentos podrán fijarse en ítems de aparente importancia dentro de agregaciones de documentos a gran escala. No es difícil imaginar futuras interfaces que podrían ofrecer a los usuarios la opción de explorar la totalidad de una gran acumulación de documentos digitales o simplemente ver la selección del archivero. De hecho, la institución archivística podría ofrecer más de una selección; se podría sugerir a aquellos usuarios interesados en un tema concreto consultar la selección A, mientras que usuarios con otros intereses podrían examinar las selecciones B, C o D. Los usuarios también podrían estar facultados para llevar a cabo sus propias selecciones y podrían publicarlas junto con la del archivero para el beneficio de otros usuarios en el futuro. Para usuarios reacios a confiar en métodos computacionales de recuperación, los sistemas en línea que ofrecen selecciones preconfiguradas proporcionarían medios alternativos para hallar material que de otra manera podría permanecer oculto a la vista en archivos muy grandes y congestionados. Aunque todas esas selecciones serían subjetivas, los archiveros podrían aplicar sus habilidades evaluativas para construirlas. Sin embargo, este enfoque diferiría de la evaluación tradicional en un aspecto crucial: no nos llevaría a la destrucción de los documentos no seleccionados. Los archivos en su totalidad seguirían estando disponibles para su inspección y análisis, y los usuarios podrían libremente hacer caso omiso de las decisiones tomadas por los profesionales.

Conclusión

Intentar predecir el futuro siempre es arriesgado y aunque casi con total seguridad la gran mayoría de los documentos se crearán en formato digital, no sabemos hasta cuándo seguiremos utilizando el papel u otros medios pre-digitales en las

próximas décadas. No podemos predecir qué nuevas formas de tecnología aparecerán o cómo afectarán a la creación y el mantenimiento de los documentos de archivo. Pero lo que sí es evidente es que los archiveros del futuro se encontrarán con cantidades sin precedentes de documentos; habrá muchos más y por lo tanto conservaremos muchos más que en el mundo analógico del pasado.

Incluso si no podemos o no tenemos la intención de conservar todos y cada uno de los documentos de la conducta humana, necesitaremos una gran transformación en la teoría y la práctica profesional si realmente queremos empezar a conservar documentos digitales a mayor escala. No solo nos obligará a dejar atrás ideas muy aceptadas sobre la necesidad de una retención altamente selectiva, sino que también afectará a la forma en la que realizamos otras funciones archivísticas como la descripción y el acceso. Debemos considerar la conservación a gran escala no como una quimera o una rendición a los valores de los entusiastas de la tecnología, sino como una oportunidad de ampliar el alcance y la utilidad de los archivos del futuro. Sin embargo, debemos reconocer que será una fuerza disruptiva que nos impulsará a desarrollar prácticas más sólidas para la llegada de un mundo de profusión digital. En la actualidad, las herramientas que necesitaremos siguen emergiendo y todavía no está muy claro cómo emplearemos estas herramientas y cómo empezaremos a tener confianza en ellas. Además, por ahora, todavía entendemos muy poco las respuestas culturales y emocionales que el uso de estas nuevas herramientas provocará. No debemos subestimar los desafíos intelectuales y prácticos que aportarán estos cambios, pero investigarlos y aceptarlos debe ser el camino a seguir para la profesión archivística ya que se enfrenta a la avalancha digital del siglo XXI.

Bibliografía

- ARORA, R., ESTEVA, M. and TRELOGAN, J. (2014) “Leveraging High Performance Computing for Managing Large and Evolving Data”, *International Journal of Digital Curation*, 9 (2), 17-27.
- BAILEY, S. (2008) *Managing the Crowd: rethinking records management for the Web 2.0 world*, Facet Publishing.
- BARON, J.R. and BORDEN, B.B. (2016) *Opening Up Dark Digital Archives through the Use of Analytics to Identify Sensitive Content*, <http://dcicblog.umd.edu/cas/wp-content/uploads/sites/13/2016/05/3.pdf>.
- BECKER, C., FARIA, L. and DURETEC, K. (2014) “Scalable Decision Support for Digital”, *OCLC Systems & Services*, 30 (4), 249-84.
- BLAIR, B.T. (2015) *Why We Are Moving to a World Where We Will Keep Everything Forever*, www.linkedin.com/pulse/why-we-moving-world-where-keep-everything-forever-barclay-t-blair.

- BOREK, A., PARLIKAD, A.K., WEBB, J. and WOODALL, P. (2014) *Total Information Risk Management*, Morgan Kaufmann.
- CANTWELL, J.D. (1991) *The Public Record Office 1838-1958*, HMSO.
- COOK, T. (1991) 'Many Are Called, but Few Are Chosen'? Appraisal guidelines for sampling and selecting case files, *Archivaria*, 32, 25-50.
- COOK, T. (1992) "Mind over Matter: towards a new theory of archival appraisal". In CRAIG, B.L. (ed.), *The Archival Imagination: essays in honour of Hugh A. Taylor* Association of Canadian Archivists.
- COX, R.J. (2004) *No Innocent Deposits: forming archives by rethinking appraisal*, Scarecrow Press.
- COX, R.J. (2011) "Appraisal and the Future of Archives in the Digital". In Hill, J. (ed.), *The Future of Archives and Recordkeeping*, Facet Publishing.
- COX, R.J. and LARSEN, R.L. (2008) "iSchools and Archival Studies", *Archival Science*, 8 (4), 307-26.
- CUKIER, K.N. and MAYER-SCHOENBERGER, V. (2013) "The Rise of Big Data: how it's changing the way we think about the world", *Foreign Affairs*, 92 (3), www.foreignaffairs.com/articles/2013-04-03/rise-big-data.
- CUNNINGHAM, A. (2008) "Digital Curation/Digital Archiving: a view from the National Archives of Australia", *American Archivist*, 71 (2), 530-43.
- DALE, K.L. (2015) "RIM's Role in Harnessing the Power of Big Data", *Information Management*, 49 (4), 29-32.
- DAVENPORT, T.H. and KIM, J. (2013) *Keeping Up with the Quants: your guide to understanding and using analytics*, Harvard Business School.
- DELSALLE, P. (1998) *Une Histoire de L'archivistique*, Presses de l'Université du Québec.
- DUCHEN, M. (1992) "The History of European Archives and the Development of the European Archival Profession", *American Archivist*, 55 (1), 14-25.
- DURANTI, L. (2015) "Archival Bond". In DURANTI, L. and FRANKS, P.C. (eds.), *Encyclopedia of Archival Science* Rowman & Littlefield.
- EIDEM, J. (2011) *The Royal Archives from Tell Leilan*, Nederlands Instituut voor het Nabije Oosten.
- ERNST AND YOUNG GLOBAL LTD (2014) *Big Data: changing the way businesses compete and operate*, [https://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/\\$FILE/EY-Insights-on-GRC-Big-data.pdf](https://www.ey.com/Publication/vwLUAssets/EY_-_Big_data:_changing_the_way_businesses_operate/$FILE/EY-Insights-on-GRC-Big-data.pdf)
- GILLILAND, A.J. (2014) "Archival Appraisal: practising on shifting sands". En BROWN, C. (ed.), *Archives and Recordkeeping: theory into practice*, Facet Publishing.
- GOLLINS, T. (2009) *Parsimonious Preservation: preventing pointless processes*, <https://www.nationalarchives.gov.uk/documents/information-management/parsimonious-preservation.pdf>
- GREENE, M.A. (2002) *Not Magic, not Science, but Art: comment on 'Archival Appraisal Alchemy' (response to Richard Cox)*, <http://web.archive.org/web/20041217002026/http://www.hfmgv.org/research/publications/symposium2002/papers/greene.asp>
- GRICKS, T.C. and AMBROGI, R.J. (2015) *A Brief History of Technology Assisted Review*, www.lawtechnologytoday.org/2015/11/history-technologyassisted-review.
- HALL, H. (1908) *Studies in English Official Historical Documents*, Cambridge University Press.
- HAM, F.G. (1993) *Selecting and Appraising Archives and Manuscripts*, Society of American Archivists.
- HARVEY, R. (2008) *Appraisal and Selection*, www.dcc.ac.uk/resources/briefingpapers/introduction-curation/appraisal-and-selection
- INFORMATION RESOURCES MANAGEMENT ASSOCIATION (2016) *Big Data: concepts, methodologies, tools, and applications*, IGI Global.
- ISAACS, E., KONRAD, A., WALENDOWSKI, A., LENNIG, T., HOLLIS, V. and WHITTAKER, S. (2013) "Echoes from the Past: how technology mediated reflection improves well-being". En CHI '13: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press.
- JENKINSON, H. (1922) *A Manual of Archive Administration*, Clarendon Press.
- JOHNSON, C. (1919) *The Care of Documents and Management of Archives*, SPCK.
- JONES, W. (2014) *Transforming Technologies to Manage Our Information: the future of personal information management*, Vol.2, Morgan and Claypool.
- KING, R., SCHMIDT, R., BECKER, C. and SCHLARB, S. (2012) "SCAPE: big data meets digital Preservation", *ERICIM News*, 89, 30-1.
- KIRSCHENBAUM, M.G. (2008) *Mechanisms: new media and the forensic imagination*, MIT Press.
- KNIGHT, J. and CUMMING, K. (2012) "Digital Data Hoarding and the Implications for RIM Professionals", *IQ: the RIM quarterly*, 28 (3), 32-5.
- KOMOROWSKI, M. (2014) *A History of Storage Cost*, www.mkomo.com/cost-per-gigabyte.
- KUMAR, V. (2017) *Big Data Facts*, www.analyticsweek.com/content/big-datafacts.
- LEAHY, E.J. and CAMERON, C.A. (1965) *Modern Records Management*, McGraw-Hill.
- LEMIEUX, V.L. (2016) "Provenance: past, present and future in interdisciplinary and multidisciplinary perspective". En LEMIEUX, V.L. (ed.), *Building Trust in Information: perspectives on the frontiers of provenance*, Springer.
- LOSEY, R. (2015) Information Governance v Search: the battle lines are redrawn, www.e-discoveryteam.com/2015/02/08/information-governance-vsearch-the-battle-lines-are-redrawn.
- MARCIANO, R. (2016) *Building a 'Computational Archival Science' Community*, www.saaers.wordpress.com/2016/07/27/building-a-computationalarchival-science-community.
- MARSHALL, C.C. (2008) "Rethinking Personal Digital Archiving Part 2: implications for services, applications, and institutions", *D-Lib Magazine*, 14 (3-4), www.dlib.org/dlib/march08/marshall/03marshall-pt2.html.
- MCGILL, A. (2016) *Can Twitter Fit Inside the Library of Congress?*, www.theatlantic.com/technology/archive/2016/08/can-twitter-fit-insidethe-library-of-congress/494339.
- MOSS, M. (2015) "What is the Same and What is". En MOSS, M. and ENDICOTT-POPOVSKY, B. (eds.) *Is Digital Different?* Facet Publishing.
- MÜLLER, O., JUNGLAS, I., VOM BROCKE, J. and DEBORTOLI, S. (2016) "Utilizing Big Data Analytics for Information Systems Research: challenges, promises and guidelines", *European Journal of Information Systems*, 25, 289-302.
- MUMMA, C.C., DINGWALL, G. AND BIGELOW, S. (2011) "A First Look at the Acquisition and Appraisal of the 2010 Olympic and Paralympic Winter Games

- Fonds: or, SELECT * FROM VANOC_Records AS archives WHERE value="true", *Archivaria*, 72, 93-122.
- NEUMAYER, R. and RAUBER, A. (2007) *Why Appraisal is not 'Utterly' Useless and Why It's Not the Way to Go Either*, www.ifs.tuwien.ac.at/~neumayer/pubs/NEU07_appraisal.pdf.
- NORA, P. (1989) "Between Memory and History: les lieux de mémoire", *Representations*, 26, 7-24.
- OPEN PRESERVATION FOUNDATION (2016) *Introducing E-ARK Specifications and the E-ARK Web Platform*, www.openpreservation.org/event/introducinge-ark-specifications-and-the-e-ark-web-platform.
- PALAIMA, T. (2003) "Archives and Scribes and Information Hierarchy in Mycenaean Greek Linear B Records". In BROSIUS, M. (ed.), *Ancient Archives and Archival Traditions*, Oxford University Press.
- RAHMAN, N. (2016) "Toward Achieving Environmental Sustainability in the Computer Industry", *International Journal of Green Computing*, 7 (1), 37-54.
- ROBERTSON, J. (2016) *E-Mail Spam Goes Artisanal*, www.bloomberg.com/news/articles/2016-01-19/e-mail-spam-goesartisanal.
- ROBYNS, M.C. (2014) *Using Functional Analysis in Archival Appraisal*, Rowman & Littlefield.
- ROSENTHAL, D.S.H. (2010) *Format Obsolescence: assessing the threat and the defences*, *Library Hi Tech*, 28 (2), 195-210.
- SHELLENBERG, T.R. (1956) *Modern Archives: principles and techniques*, F.W. Cheshire.
- SEREWICZ, L.W. (2010) "Do We Need Bigger Buckets or Better Search Engines? The challenge of unlimited storage and semantic web search for records management" *Records Management Journal*, 20 (2), 172-81.
- SHINKLE, T. (2017) "Automated Electronic Records Management: are we there yet?" *IRMS Bulletin*, 197, 14-20.
- SPENCER, R. (2017) "Binary Trees? Automatically identifying the links between born-digital Records", *Archives and Manuscripts*, 45 (2), 77-99.
- TANSELLE, G.T. (2002) "The Librarians' Double-Cross", *Raritan*, 21 (4), 245-63.
- THE NATIONAL ARCHIVES OF THE UK (2016) *The Application of Technology Assisted Review to Born-Digital Records Transfer, Inquiries and Beyond* <https://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf>
- THOMAS, D. and JOHNSON, V. (2015) "From the Library of Alexandria to the Google Campus: has the digital changed the way we do research?" En MOSS, M. and ENDICOTT-POPOVSKY, B. (eds) *Is Digital Different?* Facet Publishing.
- WADDINGTON, S., HEDGES, M., RIGA, M., MITZIAS, P., KONTOPOULOS, E., KOMPATSIARIS, I., VION-DURY, J., LAGOS, N., DARÁNYI, S., CORUBOLO, F., MULLER, C. and MCNEILL, J. (2016) *PERICLES: digital preservation through management of change in evolving ecosystems*, www.pmitzias.com/publications/PERICLES_Digital_Preservation_through_Management_of_Change_in_Evolving_Ecosystems.pdf.
- WARLAND, A. AND MOKHTAR, U.A. (2013) "Can Technology Classify Records Better than a Human?" *IRMS Bulletin*, 171, 16-19.
- WHITELAW, M. (2015) "Representing Digital Collections". In CARLIN, D. and VAUGHAN, L. (eds.) *Performing Digital: multiple perspectives on a living archive*, Ashgate.
- YEO, G. (2008) "Concepts of Record (2): prototypes and boundary objects" *American Archivist*, 71 (1), 118-43.
- YEO, G. (2012) "The Conceptual Fonds and the Physical Collection", *Archivaria*, 73, 43-80.
- YEO, G. (2013) "Trust and Context in Cyberspace", *Archives and Records*, 34 (2), 214-34.
- YEO, G. (2017) "Continuing Debates about Description". In MACNEIL, H. and EASTWOOD, T. (eds.) *Currents of Archival Thinking, 2nd ed.*, Libraries Unlimited.