

Retención de la información en la empresa: cuestión de sostenibilidad

Information retention in companies: a question of sustainability

Elisa García-Morales

García-Morales, Elisa (2021). "Retención de la información en la empresa: cuestión de sostenibilidad". *Anuario ThinkEPI*, v. 15, e15f03.

<https://doi.org/10.3145/thinkepi.2021.e15f03>

Publicado en *IweTel* el 30 de noviembre de 2021

Elisa García-Morales

<https://orcid.org/0000-0003-2675-3046>

Inforarea SL

garcia-morales@inforarea.com



Resumen: Guardar toda la información indefinidamente no es sostenible. La retención es un aspecto esencial de la gobernanza de la información que facilita la gestión del ciclo de vida de datos y documentos. Las tendencias futuras de las tecnologías de información nos llevan a reflexionar sobre las metodologías disponibles y las cualificaciones profesionales necesarias para encontrar nuevos caminos que ayuden la definición y aplicación de las necesarias políticas de retención en empresas y organizaciones.

Palabras clave: Retención de información; Retención de documentos; Gobierno de información; Sostenibilidad; Ciencia archivística computacional; Tecnologías emergentes.

Abstract: It is not sustainable to keep all information indefinitely. Retention is an essential aspect of corporate information governance that facilitates data and document lifecycle management. Future trends in information technologies lead us to reflect on the methodologies available and the professional qualifications needed to find new ways to help define and implement the necessary retention policies in companies and organizations.

Keywords: Information retention; Records retention; Information governance; Sustainability; Emerging technologies; Computational archival science.

1. Necesidad de un programa de retención de la información

La retención de información es el proceso que permite:

- determinar qué información debe conservar una empresa u organización y por cuánto tiempo: esto se concreta, a partir del análisis y valoración previa, en unos instrumentos denominados tablas de retención;
- qué se debe hacer cuando cumple el período de retención determinado (disposición): dichas acciones deben ser aprobadas por las instancias con autoridad suficiente sobre la información;
- ejecutar la disposición, que consiste básicamente en la eliminación de información o su transferencia a estados/entornos de conservación a largo plazo: ello requiere que la información disponga de algún tipo de mecanismo de identificación, clasificación y/o catalogación (metadatos) para poder aplicar lo dispuesto.

El término retención se emplea en el contexto del "*records management*" y/o del gobierno de la información. En su esencia y bases metodológicas, adopta los conceptos de la disciplina archivística para establecer los calendarios de conservación de documentos. Estos son instrumentos de uso extendido

en entidades y administraciones públicas obligadas por ley a conservar información y documentos. En el sector privado, el concepto “retención” tiene un significado más enfocado a facilitar la eliminación controlada de información que a su conservación para el futuro.

La empresa actual produce una gran cantidad de documentos y datos que constituyen evidencia de su actividad y son garantía frente a posibles litigios y también fuente esencial para el análisis, predicción y toma de decisiones. Por otro lado, también se produce demasiada información ROT (Redundante, Obsoleta o Transitoria) que no tiene ningún valor para la organización en el medio largo plazo. La intensificación de la digitalización provoca el crecimiento continuo de la información electrónica, que aún sigue coexistiendo con documentos en papel. Guardarlo todo indefinidamente –como algunos preconizan– no es sostenible y “sostenibilidad” es hoy en día una palabra clave en el gobierno empresarial. Algunos de los argumentos tradicionales que justifican un programa de retención de la información son:

- elevado coste de almacenar la documentación y la información (física y electrónica) indefinidamente;
- riesgos de mantener en el tiempo información obsoleta que pudiera estar sujeta a escrutinio de terceros;
- necesidad de hacer defendible ante los tribunales la retención y la destrucción de documentos o datos mediante unos protocolos de actuación fundamentados;
- existencia de requerimientos legales para mantener determinados registros durante períodos específicos de tiempo.

Pero la realidad es que sobre este tema “se habla mucho y se aplica poco”. Hay coincidencia amplia sobre la importancia de la gestión del ciclo de vida de la información, sobre la necesidad de cumplir la normativa, en especial la que afecta a la protección de datos y sobre la conveniencia de establecer políticas de retención. Cada vez es más frecuente que en las memorias y/o códigos de conducta de las grandes compañías encontremos referencias a la retención de información. Sin embargo, en la empresa privada, los casos de uso son muy escasos. No ocurre así en el ámbito público donde es más fácil encontrar calendarios de conservación de documentos publicados. En mi experiencia he observado que la mayoría de las empresas desconocen y ni se han planteado elaborar unas tablas de retención y que, cuando existen, quedan en meros ejercicios teóricos: bien porque no se aprueban por las instancias competentes, o bien porque la disposición no se ejecuta, o se ejecuta solo a conjuntos muy parciales de documentos.

¿Ocurre esto por desinterés de los directivos?, ¿por la renuencia a asumir responsabilidades sobre la destrucción de la información?, ¿por la resistencia de los usuarios finales?, ¿por el miedo a que los datos eliminados pudieran servir para futuros procesos de análisis?, ¿o por la complejidad de las metodologías y procesos para aplicar la retención? Posiblemente hay una combinación de todas estas razones, unida a la sensación de inabarcabilidad que produce el crecimiento continuo de la información. Las tendencias en el desarrollo de la tecnología que se apuntan ya hoy y marcarán los próximos años, nos debieran hacer reflexionar en qué medida los planteamientos actuales en relación con la retención de información ofrecen una solución viable y práctica para las necesidades empresariales.

“La retención de información es el proceso que permite determinar qué información debe conservar una empresa u organización y por cuánto tiempo; qué se debe hacer cuando cumple el período de retención determinado y ejecutar la disposición, que consiste en la eliminación de información o su transferencia a estados/entornos de conservación a largo plazo”

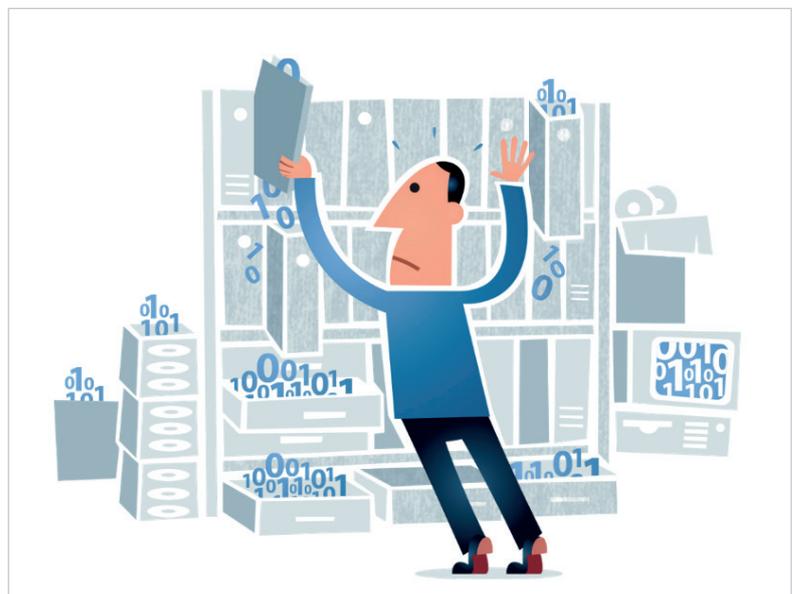


Illustration by Jørgen Stamp. Digitalbevaring.dk

2. Evolución de la tecnología e implicaciones en la retención de la información

Algunas de las tendencias que apuntan los expertos para la evolución de las tecnologías de información en los próximos años son:

- Continuará el crecimiento imparable de la información o la “data” en su sentido holístico, lo que incluye cada vez mayor variedad de tipos y formatos. Esta tendencia afecta al núcleo del QUÉ es lo que debiera estar sujeto a los procesos de retención. Hasta ahora la metodología, procesos y programas de retención se han focalizado sobre los documentos en su forma más tradicional –papel o electrónica–; el futuro nos exige que estas metodologías sean aplicables a datos estructurados y no estructurados de muy diversas características. Los términos retención de documentos e incluso “*records retention*” se perciben como algo limitado al documento en sus formas más tradicionales. Debemos hablar y pensar en términos data/información como ya proponen algunos profesionales pioneros (Casellas, 2019). Por otro lado, el potencial valor de los datos para operaciones analíticas y para el aprendizaje de máquinas debe ser contemplado como un elemento añadido en las tareas de valoración para establecer las tablas de retención (Barragan, 2019).
- La información tenderá a originarse de manera más descentralizada y cada vez más lejos de los sistemas transaccionales, como ya ocurre actualmente. La generación de información cada vez está más dominada por lo que se denomina “*non-data sources*” o fuentes de origen que no son datos estructurados. Los dispositivos externos son capaces de recolectar, procesar y almacenar información por sí mismos, sin necesidad de conectar con redes o infraestructuras centralizadas de almacenamiento. El “*edge computing*” como concepto que traslada el proceso de información a los dispositivos propiamente dichos en busca de una mayor agilidad, es otra de las tendencias que se impone; ello puede suponer el traslado de los requerimientos de retención hacia miles de dispositivos externos con el consiguiente aumento de la complejidad de su control y gestión. Uno de los fundamentos de la archivística, la gestión documental y la retención reside en el denominado “principio de procedencia” (Ainy et al., 2015). En este escenario futuro nos preguntamos ¿QUIÉN origina la información que se debe retener? (Buneman; Tan, 2018). Ya no son suficientes las entidades o agentes tradicionales procedentes de un mundo centralizado, físico, con estructuras definidas y más o menos jerarquizadas. El principio de procedencia es esencial para entender el contexto en el que la información se produce, el posible grado de responsabilidad del productor y también para determinar qué se debe de hacer con ella en el tiempo.
- Se acentuará la tendencia al almacenamiento de la información en la nube. El crecimiento de la “data” va a hacer que las organizaciones busquen cada vez más soluciones de almacenamiento en la nube o soluciones de tipo híbrido, que a día de hoy prometen capacidades prácticamente “ilimitadas”. Parecería que esta es la solución mágica para despreocuparse de la necesidad de eliminar información, pero todo apunta a que esta hiperacumulación de información conlleva unos costes importantes y no es sostenible. Hasta ahora, la retención podía ser más fácilmente aplicable puesto que la mayoría de las empresas disponían de su propia infraestructura de almacenamiento centralizada y segura. Esta dinámica hacia el *cloud computing* traslada a los proveedores de infraestructuras la responsabilidad sobre la operación y el mantenimiento del almacenamiento. Se complica por tanto el DÓNDE aplicar la retención, que puede quedar en muchos casos sujeta a controles, reglas o restricciones de los propios proveedores. Tampoco se debe desdeñar el riesgo de su desaparición de mercado, como ya ha ocurrido en numerosas ocasiones.
- Cambios en la arquitectura para la gestión masiva de datos. Y con ello nos referimos a la expansión de los “*data lakes*” como los nuevos entornos en los que se cocinan las grandes masas de información para su explotación analítica. Estamos hablando de datos en bruto, que proceden –son copiados o duplicados– de otras fuentes o bases de datos, almacenados en una arquitectura plana muy diferente a las entidades que hasta ahora considerábamos a efectos de la retención. Cómo conseguir que estos lagos no se conviertan en ciénagas de información está estrechamente relacionado con la retención (Shacklett, 2017). Nuevamente nos enfrentamos al reto de definir QUÉ información de un *data lake* debiera ser objeto de la retención, su relación con la información origen, CÓMO establecer categorías para aplicar una retención y CÓMO ejecutar la disposición.

“Guardarlo todo indefinidamente –como algunos preconizan– no es sostenible y ‘sostenibilidad’ es hoy en día una palabra clave en el gobierno empresarial”

- Los sistemas de aprendizaje automático e inteligencia de máquina se van a imponer en todos los ámbitos permitiendo dotarnos de capacidades avanzadas de análisis y proceso. Serán los futuros generadores de una parte importante de la información: todo un nuevo mundo que va a requerir nuevas capacidades para valorar qué y cuándo se debe eliminar o conservar.

3. Algunas conclusiones sobre las fortalezas y debilidades de las metodologías de retención

La propuesta metodológica de la archivística y gestión documental para establecer y aplicar la retención tiene el punto fuerte de ser fruto de años de trabajo y experiencia. Pero sus fundamentos teóricos y prácticos se desarrollaron para un contexto analógico mucho más limitado y abarcable y ello es un condicionante para la aplicación práctica de los programas de retención. Los líderes del mercado informático rápidamente empiezan a tomar posiciones cuando perciben que la gestión del ciclo de vida es un factor importante para el negocio. Sin ir más lejos véase el desarrollo de las funcionalidades de “compliance” en el entorno *Microsoft O365*. Ahora también entra *Amazon* (Kollarath, 2020) en el mundo de la retención de datos y otros actores del abordan la cuestión desde la perspectiva del *e-discovery* para el ámbito jurídico.

A pesar de los meritorios esfuerzos de algunos profesionales de la información por evolucionar la disciplina para adaptarse a las nuevas circunstancias del mundo “data”, resulta difícil salir del reducto histórico del archivo y posicionarse en el núcleo tecnológico donde se producen los cambios en el escenario de la información digital.

Las bases teóricas y metodológicas desarrolladas en el campo de la archivística y la gestión documental son un excelente punto de partida, pero resultan insuficientes para el contexto tecnológico actual y requieren revisión continuada a la luz de las nuevas situaciones. Una de las aportaciones más valiosas desde el punto de vista práctico, ha sido el enfoque metodológico de *Big Buckets*¹, que facilita tanto el diseño como la implementación de grandes contenedores de clasificación para la aplicación de los períodos de retención. Este fue propuesto y desarrollado por *NARA*, la *Administración de Archivos y Registros Nacionales* de Estados Unidos, como estrategia para mejorar la efectividad en la retención de registros de las agencias federales. Hoy en día ha sido ampliamente adoptado en el mundo anglosajón tanto por organismos públicos como empresas privadas (Saffady, 2018). No podemos decir lo mismo en el ámbito castellanohablante donde apenas encontramos referencias al uso de esta metodología. Las investigaciones del grupo *Interpares Trust Project* sobre la confiabilidad de los documentos en la nube y otros entornos de computación distribuida, así como la problemática de su conservación, apuntan planteamientos interesantes en relación con la retención (Franks, 2018). Las aportaciones de autores como Marciano y Lemieux sobre el *blockchain* y el *big data* son valiosísimas. En otros terrenos, proyectos como *Archiver* de la UE² para la preservación de datos de investigación pueden ayudar a comprender mejor cómo manejar grandes cantidades de datos estructurados para su retención a largo plazo. La aplicación de la inteligencia artificial, o mejor dicho el aprendizaje de máquinas es el camino futuro. Se deben investigar los fundamentos lógicos, algorítmicos y metodológicos necesarios para el desarrollo de aplicaciones para la retención y ello requiere perfiles con experiencia tecnológica.

Pero leyendo lo que se escribe desde distintos ámbitos –TIC, records managers, expertos legales– se llega a la conclusión de que la comunicación “interprofesional” es muy escasa y las reflexiones y los hallazgos no acaban de conectar con una realidad empresarial que exige soluciones poco intrusivas. Ante la evolución de las técnicas existentes y las tendencias TI que se perfilan, la intercomunicación entre ambos campos resulta imprescindible. En este sentido, el planteamiento de evolucionar hacia una disciplina de *Computational Archival Science* (Marciano, et al., 2018) resulta un enfoque de futuro realmente interesante. No será posible hallar nuevos caminos sin el uso intensivo de la tecnología y la innovación. En definitiva, o bien los profesionales de la información participan activamente en las etapas de diseño TI, o bien los fundamentos teóricos de la ciencia archivística permean en la formación de los tecnólogos.

“El principio de procedencia es esencial para entender el contexto en el que la información se produce, el posible grado de responsabilidad del productor y también para determinar qué se debe de hacer con ella en el tiempo”

4. Notas

1. ARMA resources: *Global retention schedule challenges and choices*.
<https://www.arma.org/store/viewproduct.aspx?ID=10539972>
2. Archiver-project. *Archiving and preservation for research environments*.
<https://archiver-project.eu>

5. Referencias

- Ainy, Eleanor; Bourhis, Pierre; Davidson, Susan B.; Deutch, Daniel; Milo, Tova** (2015). "Approximated summarization of data provenance". In: *Proceedings of the CIKM*, pp. 483-492.
<https://doi.org/10.1145/2806416.2806429>
- Barragan, Salvador P.** (2019). "Information governance maturity model: Should retention be rethought?". In: Katuu, Shadrack (ed.). *Diverse applications and transferability of maturity models*. IGI Global Publisher. ISBN: 978 1522570806
<https://doi.org/10.4018/978-1-5225-7080-6.ch004>
- Buneman, Peter; Tan, Wang-Chiew** (2018). "Data provenance: What next?". *ACM SIGMOD record*, v. 4, n. 3, pp. 5-16.
<https://doi.org/10.1145/3316416.3316418>
- Casellas, Lluís-Esteve** (2019). "Archivos, documentos, datos. De la gestión a la gobernanza". En: *VI Congreso de Innovación Pública Novagob*. Sevilla, 16-18 octubre.
<https://www.archiverosdeandalucia.org/wp-content/uploads/2019/11/Ponencia-Lluis-Esteve-Casellas-Novagob-2019.pdf>
- Franks, Patricia C.** (2018). "Trusting records and data in the cloud: The creation, management, and preservation of trustworthy digital content". In: *Encyclopedia of archival science*. University of British Columbia, Vancouver: Facet, pp. 117-134.
<https://doi.org/10.29085/9781783304042.006>
- Kollarath, Prasanth** (2020). *Building data lakes and implementing data retention policies with Amazon RDS snapshot export to Amazon S3*, 29 may.
<https://aws.amazon.com/es/blogs/database/building-data-lakes-and-implementing-data-retention-policies-with-amazon-rds-snapshot-export-to-amazon-s3>
- Marciano, Richard; Lemieux, Victoria; Hedges, Mark; Esteva, Maria; Underwood, William; Kurtz, Michael; Conrad, Mark** (2018). "Archival records and training in the age of big data". In: Percell, Johna; Sarin, Lindsay C.; Jaeger, Paul T.; Bertot, John-Carlo (eds.). *Re-envisioning the MLS: Perspectives on the future of library and information science education*. ISBN: 978 1 78754 885 5
- Saffady, William** (2018). *Records management experience with big bucket retention: A status report*. ARMA International Education Foundation.
<http://armaedfoundation.org/wp-content/uploads/2018/08/AIEF-Research-Paper-Saffady-Big-Buckets-2018-081518.pdf>
- Shacklett, Mary** (2017). "How to keep your data lakes from becoming cesspools". *TechRepublic*, 23 May.
<https://www.techrepublic.com/article/how-to-keep-your-data-lakes-from-becoming-cesspools/>



<http://www.profesionaldelainformacion.com>

Bienvenido a EPI
Revista científica internacional

e-ISSN: 1699-2407
<https://doi.org/10.3145/EPI>

Revista internacional de
Información y Comunicación
Indexada por WoS Social Sciences Citation Index (Q3),
Scopus (Q1) y otras bases de datos

Factor de impacto JCR:
JIF 2020=2,253

Scopus/SCImago Journal Rank:
SJR 2020=0,698