



DOI: 10.11565/GESTEN.V4I2.89



Preservación Digital



Claudio Escobar Arriagada

Master in Business Engineering
(MBE), U. de Chile
Académico FEN-UAH
clescoba@uahurtado.cl

RESUMEN

La preservación digital es un conjunto de tareas, técnicas, métodos y tecnologías que buscan garantizar el acceso a la información en formato digital. Este artículo presenta una aproximación a los conceptos clave relacionados con la preservación digital, los beneficios de su implementación, los principales aspectos tecnológicos asociados, los desafíos y una mirada a la situación actual de la preservación digital en Chile.

Preservación digital: tecnología y desafíos

Desde finales de los años 80 y comienzos de los 90 comenzó la masificación del uso de computadores personales tanto en el mundo laboral como particular. Este hecho ha provocado, en primer lugar, que cada vez se generen más y más documentos (cada dos años se duplica la cantidad de información accesible en internet (International Data Corporation, 2011)). En segundo lugar, con el correr de los años se ha observado un fenómeno que se repite cada vez con mayor rapidez, la obsolescencia tecnológica, la cual se puede definir como la necesidad de recambio de un dispositivo tecnológi-

co debido principalmente a la aparición de nuevas tecnologías (más eficientes), imposibilidad de encontrar repuestos, o simplemente por estrategias comerciales. Un claro ejemplo de ello lo podemos observar en la extinta hoja de cálculos Lotus 1-2-3 que llegó a ser el software más utilizado por las empresas hasta finales de los años 80, pero que dejó de existir hace aproximadamente 20 años. En base a lo anterior surge la siguiente interrogante ¿Cómo podemos garantizar el acceso a la información contenida en documentos que han sido generados con software que en unos años quedará obsoleto o almacenados en soportes que no estarán vigentes en poco tiempo más? La preservación digital ofrece una respuesta a esta pregunta.

¿Qué es la preservación digital?

La preservación digital tiene por objetivo principal el aseguramiento del acceso a los contenidos de documentos electrónicos¹. Según Jones y Beagrie (2001) la preservación digital comprende una serie de tareas y actividades necesarias, planificadas y administradas para asegurar el acceso continuo a materiales digitales por el periodo que sea necesario. Se refiere a todas las acciones requeridas para mantener el acceso a los materiales digitales aún después de que se presenten fallas en los medios de almacenamiento o haya cambios tecnológicos. Los autores plantean que la preservación digital se puede clasificar en:

¹ Para un diagnóstico del estado de la gestión documental en instituciones públicas, ver Microsystem, 2016.

Preservación de corta duración: Busca garantizar el acceso al contenido de objetos digitales por un periodo de tiempo definido, o bien, que el uso de los documentos sea calculado para un periodo de tiempo menor al de los cambios tecnológicos.

Preservación de duración media: Busca garantizar el acceso al contenido de objetos digitales aún después de que ocurran cambios tecnológicos, o bien, dentro de un periodo de tiempo definido, pero no indefinidamente.

Preservación de larga duración: Busca garantizar el acceso al contenido de objetos digitales de forma indefinida.

Cada organización puede planificar y realizar las acciones que estime convenientes para preservar digitalmente sus activos digitales, pero existen modelos específicos para ello. Uno de los más utilizados es el modelo de referencia OAIS (Open Archival Information System), el cual a partir del año 2003 pasó a ser una norma ISO (14.721). Esta norma, revisada y actualizada el año 2012, da los lineamientos generales respecto a las acciones que se deberían realizar en una organización para preservar digitalmente sus documentos, pero da la libertad de determinar la mejor forma de llevarlas a cabo y con la tecnología que cada organización prefiera.

La Norma propone un modelo de referencia que busca proporcionar una comprensión más amplia de lo que se necesita para preservar la información y acceder a ella a largo plazo, teniendo en consideración aspectos como el avance de la informática y el crecimiento exponencial de información en formatos digitales. En otros términos, OAIS viene a ser una entidad cuyo interés radica en conservar la información que se genera al interior de una organización, a fin de que uno o múltiples grupos de usuarios específicos, que deberían ser capaces de comprender el conjunto particular de información que allí se resguarda (comunidad específica), puedan acceder y usar dicha información, cada vez que lo requieran.

En este sentido, se plantea como necesidad que la información custodiada esté sujeta al proceso de preservación a largo plazo; es decir, al acto de mantener la in-

formación comprensible de forma independiente, con rasgos que den garantía de su autenticidad, por un periodo de tiempo lo suficientemente grande como para verse afectada por cambios, tanto, en lo tecnológico como en la comunidad específica a la que sirven.

En la preservación digital, bajo la mirada del modelo OAIS, se distinguen principalmente 3 tipos de actores:

Productores: Personas o sistemas que proporcionan o generan la información que se desea o requiere preservar.

Consumidor: Personas o aplicaciones que interactúan con los servicios del sistema para, fundamentalmente, localizar y recuperar la información preservada de acuerdo con su interés.

Dirección: Corresponde a las entidades que establecen la política general del sistema como parte de una política más amplia. Es decir, busca que las políticas de trabajo del sistema vayan en armonía con las políticas generales del entorno donde se encuentra inserto (por ejemplo, una organización, entidad gubernamental, entre otros).

Además de los actores, el modelo OAIS se refiere a paquetes de información. Estos paquetes de información, a grandes rasgos, corresponden a la unión de objetos de datos de contenido (documentos) y su información de representación asociada (metadata), lo que quiere decir que para preservar documentos digitales no solamente es necesario almacenar los archivos, sino que además éstos deben ser previamente clasificados y descritos. La clasificación es fundamental para que los documentos preservados conserven el contexto en el que fueron creados y entender su procedencia, mientras que la descripción es necesaria para obtener información adicional a la que contienen los documentos y que eventualmente serán utilizados para realizar búsquedas. Para la realización de la descripción exis-

ten varios esquemas de metadatos que se pueden emplear, entre los que destacan Dublin Core, EAD y METS. Los documentos que son enviados al sistema de preservación se empaquetan y dan forma a un Paquete de Información de Transferencia (PIT), los cuales a la hora de ser preservados en el sistema se les agrega nueva información (útil para la preservación a largo plazo) dando origen a un Paquete de Información de Archivo (PIA), que finalmente darán origen a los Paquetes de Información de Consulta (PIC), que son los documentos que finalmente podrán ser usados y consultados por los usuarios finales del sistema. Para empaquetar los documentos y su metadata se debe estructurar los contenidos, para ello existen diversas especificaciones, entre las que destacan BagIt y E-ARK.

“Preservar documentos digitales no solo implica el desarrollo y/o implementación de uno o más sistemas, como se mencionó anteriormente, se requiere de un trabajo previo en la organización donde se implementará el sistema”

La mayor cantidad de tareas se realiza a la hora de crear un Paquete de Información de Archivos, que es el paquete que se preservará a largo plazo. Las principales tareas que se realizarán sobre este paquete de información abordan chequeos de integridad, presencia de virus, control de accesos y, sobre todo la realización de acciones orientadas a la preservación. Existe más de una estrategia para preservar documentos digitales, siendo las más utilizadas la emulación que consiste en desarrollar y emplear un software que permita acceder al contenido de un documento en un formato obsoleto y la migración que requiere de ir generando nuevos archivos en formatos de preservación (generalmente en formatos libres, no comerciales)

Tecnología

Como se puede observar, las tareas asociadas a la preservación digital son numerosas y algunas de ellas complejas, por lo que el sistema por implementar en una organización que preservará digitalmente sus documentos debe ser lo suficientemente robusta para dar respuesta a todas esas acciones y ser sustentable

“En la actualidad existen soluciones informáticas de tipo comercial y open source desarrolladas específicamente para abordar la preservación digital y que poseen más de 10 años en el mercado”

en el tiempo. En ese contexto se deben considerar las siguientes recomendaciones a la hora de desarrollar o implementar un sistema:

Open Source: Preferir soluciones de código abierto, con una amplia comunidad de desarrolladores y de mantenimiento sencillo que no impliquen escalamiento de costos en el tiempo.

Multiplataforma: Permitir el trabajo sobre diversos sistemas operativos, en particular Linux (en sus diversas distribuciones), Windows, entre otros. El sistema, al igual que los documentos a preservar, debe estar pensado sin una fecha de término, por lo que esta característica es fundamental para poder realizar futuras migraciones.

Escalabilidad: Contar con la habilidad de reaccionar y adaptarse a altas cargas de trabajo sin perder su rendimiento (performance).

Seguridad: Contar con altos estándares de seguridad en las siguientes áreas: integridad de la información transferida, acceso controlado, bitácora de trazabilidad, manejo de perfiles y roles de usuarios, alertas operativas y todo otro elemento requerido para garantizar la información transmitida y almacenada.

Mantenibilidad: Contar con herramientas que permitan realizar mantenimiento a nivel usuario, es decir, que no requiera de la concurrencia de un programador para hacer cambios de configuración básicos.

Capacidad de recuperación: El sistema debe ser estable desde el punto de vista operativo, con mecanismos de recuperación en caso de caídas y/o fallas.

Modularidad: El sistema completo debe ser de carácter modular. Se recomienda el uso de microservicios en vez del desarrollo monolítico (modelo clásico en el desarrollo de software), es decir, que cada funcionalidad o grupo de funcionalidades se desarrolle como un componente, los que deberán estar desarrollados bajo tecnologías abiertas y conocidas. Esta característica permitirá, en el futuro, poder realizar actualizaciones y/o mejoras de una manera rápida, sencilla y menos costosa.

Uso de formatos abiertos: Las estrategias para preservar los documentos digitales se basan principalmente en los formatos de los documentos. Se recomienda fuertemente el uso de formatos libres para generar los archivos de preservación y acceso, principalmente para no generar dependencia de un software y/o proveedor en particular.

Tabla 1. Equivalencia entre formatos cerrados y libres

Fuente: La preservación, otra ventaja del software libre y de código abierto

Tipo de formato	Formato cerrado	Formato abierto
Documento de texto	DOC, DOCX	ODT, PDF
Hoja de cálculo	XLS	ODS
Archivo de audio con pérdida	MP3, AAC	OGG
Archivo de audio sin pérdida	WAV	FLAC
Compresión de archivos	ZIP, RAR	7Z

En la actualidad existen soluciones informáticas de tipo comercial y open source desarrolladas específicamente para abordar la preservación digital y que poseen más de 10 años en el mercado. El funcionamiento de dichos softwares está basado en el modelo OAIS y trabajan en base a los estándares y especificaciones mencionados con anterioridad. En el ámbito comercial destaca el software Libsafe², mientras que en el ámbito open source destacan los softwares RODA³ y Archivematica⁴.

Principales desafíos

Preservar documentos digitales no solo implica el desarrollo y/o implementación de uno o más sistemas, como se mencionó anteriormente, se requiere de un trabajo previo en la organización donde se implementará el sistema. Por una parte, se debe realizar un trabajo de carácter más técnico en el cual se realice un diagnóstico y análisis de la organización y así determinar qué documentos se deberán preservar, realizar un trabajo exhaustivo para clasificar esos documentos y determinar el nivel de detalle con el que se describirán los documentos. Por otro lado, es necesario realizar un trabajo orientado a preparar la organización para que desarrolle las capacidades necesarias asociadas a la preservación digital. En ese sentido, se hace necesario contar con el apoyo del cuerpo directivo de la organización, y en segundo lugar el trabajo con las personas que estarán directamente relacionadas con las actividades a realizar. Los principales desafíos a tener en consideración son:

- Obtener el respaldo de las autoridades de la organización
- Contar con un equipo de personas con los conocimientos idóneos para realizar el trabajo, entre los que destacan: gestión de procesos, normas de descripción archivísticas, gestión documental, esquemas de metadatos, programación, modelo OAIS, entre otras.
- Contar con un buen plan de gestión del cambio
- Desarrollar una cultura orientada a la gestión de los documentos, priorizando cada vez más el uso de los documentos digitales

² <https://www.libnova.com/preservacion-digital-libsafe-y-libdata/>

³ <https://demo.roda-community.org/#welcome>

⁴ <https://www.archivematica.org/es/>



- Asegurar un presupuesto permanente en el tiempo destinado a estas actividades
- En muchos casos se hace necesario convencer al área informática para poder usar software open source
- Definir estrategia tecnológica respecto al almacenamiento ¿Preferir arquitectura on-premise o cloud?
- Realizar vigilancia tecnológica de forma constante, sobre todo en lo relacionado con preservación digital.

Preservación digital en Chile

En la región, y específicamente en Chile, la preservación digital es algo relativamente incipiente, no existiendo casos en los que se pueda apreciar la implementación completa del modelo OAIS. Sin embargo, el Archivo Nacional de Chile está en pleno desarrollo de un proyecto que

busca, entre otras cosas, implementar un sistema que le permita recibir transferencias de documentos electrónicos desde los organismos públicos para almacenarlos y preservarlos a largo plazo. El proyecto se basa en la implementación del modelo OAIS, por lo que se hizo necesario el desarrollo de software especializado para cada una de las etapas del modelo de Información de Archivos, en la que se empleó el software de preservación digital Archivematica. Muchos de los desafíos mencionados están o han estado presentes en el desarrollo del proyecto, algunos de ellos aún están en una etapa de definición, mientras que otros ya fueron resueltos como el uso de tecnología open source, en la guía Lineamientos para desarrollo de software (Gobierno Digital, 2018) se hace hincapié en el uso de este tipo de software por sobre el comercial.

Reflexiones finales

La preservación digital es fundamental para resguardar el patrimonio de las organizaciones y en ese sentido no basta con la realización de respaldos de documentos. El aseguramiento de que a largo plazo se pueda acceder a los contenidos de documentos debe ser una preocupación real y abordada seriamente, algo que hoy no ocurre y que requerirá de un gran esfuerzo, pues la brecha para lograrlo es muy grande. Se requiere de voluntades, cambio de cultura organizacional, recursos y de personal con los conocimientos y competencias necesarias para lograrlo. El proyecto de modernización del Archivo Nacional es un primer esfuerzo por lograrlo y podría servir de referencia para otras organizaciones que se quieran sumar a la preservación digital.

Bibliografía

- Castellano, G. (s.f.). La preservación, otra ventaja del software libre y de código abierto. Obtenido de Open Expo Europe: <https://openexpo.europe.com/es/preservacion-software-libre-codigo-abierto/>
- Council of the Consultative Committee for Space Data Systems. (2012). Reference Model for an Open Archival Information System (OAIS). Washington, DC.
- Gobierno Digital. (2018). Lineamientos para desarrollo de software: guía técnica.
- International Data Corporation (IDC). (2011). The 2011 Digital Universe Study: extracting value from chaos. EMC.
- Jones, M., & Beagrie, N. (2001). Preservation Management of Digital Materials. British Library Cataloging in Publication Data.