

# **PRESERVACIÓN DIGITAL PRÁCTICA: EL USO DE LAS HERRAMIENTAS LIBSAFE Y LIBDATA**

EFRAIM MARTÍNEZ

# PRESERVACIÓN DIGITAL: RETOS Y PROPUESTAS ACTUALES

## Preservación digital práctica: el uso de las herramientas Libsafe y Libdata

---

EFRAIM MARTÍNEZ GUTIÉRREZ  
*Responsable de marketing y estrategia*  
efraim.martinez@libnova.es

### ÍNDICE DE CONTENIDOS

1. Introducción
2. Qué es la preservación digital
3. Breves notas sobre la preservación digital a largo plazo
4. ¿Por qué hacer preservación digital?
5. Pasos y fases de la preservación digital
6. Libsafe: Herramienta de preservación digital a largo plazo
7. Libdata: Almacenamiento diseñado para preservación digital
8. Organizando un plan de preservación con Libsafe y Libdata
9. Conclusiones

## 1. INTRODUCCIÓN

Cada vez más el conocimiento se genera, convierte y almacena en formato digital. Por ello, todas las organizaciones, tanto públicas como privadas, deben tomar medidas para garantizar la conservación y reutilización de este conocimiento para el futuro.

Por un lado percibimos la evolución de la necesidad de preservación, que ha comenzado por aquellas instituciones más cercanas al mundo de los bienes culturales e históricos (las que ya la ejercían sobre artefactos analógicos) y está introduciéndose en otros sectores hasta ahora mucho menos sensibles.

Además el mundo digital supone una diferencia cualitativa que hace que las herramientas, procesos y actores involucrados sean distintos de los que estábamos acostumbrados.

Bien sea por coste, por falta de recursos humanos o por complejidad, aún pocos organismos se han decidido a comenzar a estudiar este aspecto, lo que puede conllevar el riesgo en los próximos años de grandes pérdidas de información.

Libnova, mediante el desarrollo de *Libsafe* y *Libdata*, propone un nuevo enfoque para abordar su proyecto de preservación digital con recursos económicos accesibles mediante la aplicación de metodologías estándar y con herramientas sencillas y fáciles de implementar.

Recorreré en este texto algunos de los problemas prácticos a los que se enfrentan las Instituciones cuando plantean un proyecto de preservación digital y la forma como proponemos resolverlos con *Libsafe* y *Libdata*.

## 2. QUÉ ES LA PRESERVACIÓN DIGITAL

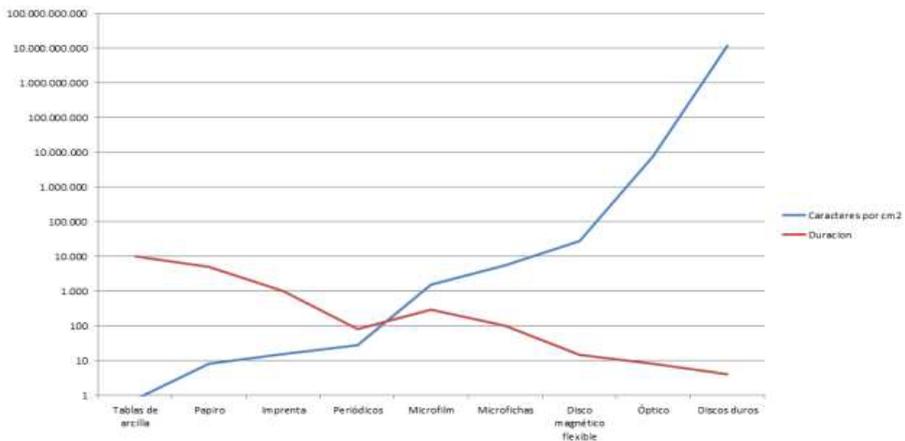
Definimos la **preservación digital a largo plazo** como el conjunto de actividades que se realizan sobre la información digital que aseguran el acceso, integridad, autenticidad y reutilización de la misma en el futuro.

Es, por lo tanto, una tarea continua en el tiempo, que se inicia con la creación de la información u objetos digitales y les acompaña a lo largo de todo su ciclo de vida, y que está centrada en los usuarios futuros y los usos que estos harán de la información.

## 3. BREVES NOTAS SOBRE LA PRESERVACIÓN DIGITAL A LARGO PLAZO

La codificación digital presenta numerosas ventajas, como facilidad de manipulación y densidad de almacenamiento, pero, a diferencia de lo que ha ocurrido con los soportes tradicionales, no podemos confiar en la suerte o en la “*negligencia benigna*” cuando pretendemos hacer que la información digital perdure en el tiempo. Toda la información digital sobre la que no se toman acciones concretas para su preservación desaparece en un corto periodo de tiempo. Veamos los motivos:

- En primer lugar, la información generada digitalmente y digitalizada es mucho más extensa, lo que hace más compleja su selección y más costoso su mantenimiento.
- Por otro lado, en el mundo digital soporte y contenido se separan: mantener y conservar el soporte no garantiza la conservación del contenido, y mucho menos su usabilidad.
- Los soportes son, además, mucho menos duraderos. En la siguiente figura puede observarse la densidad de almacenamiento por centímetro cuadrado frente a la duración estimada en años. Teniendo en cuenta que la escala lateral es logarítmica, actualmente hay 10 decenas de billones más densidad de almacenamiento que en los comienzos de la escritura.



- Además, tanto los contenidos como los soportes están sujetos a “obsolescencia tecnológica”. Para preservar conocimiento en formato digital debemos pensar en los medios de lectura y acceso, además de manuales, procedimientos de operación y muchos más detalles.

#### 4. ¿POR QUÉ HACER PRESERVACIÓN DIGITAL?

Ya que hemos visto que la preservación es una tarea compleja y costosa, procede en este momento reflexionar sobre los motivos que nos llevan a abordarla:

- En algunos casos, los motivos son **culturales e históricos**. Bien porque el propio bien digital sea en sí mismo preservable o bien porque, contando con él, somos capaces de mantener en mejores condiciones el artefacto original.
- A veces, el **valor económico de la información** es el principal motivador. Sea por la dificultad de volver a obtenerla (pensemos en los resultados de

las misiones espaciales) como por su uso en el futuro (como en el caso de los datos geológicos que se almacenan para ser usados más adelante con nuevos algoritmos en busca de yacimientos).

- Quizás es un **motivo legal** el que nos impulsa a preservar el dato, si algún tipo de entidad reguladora o normativa nos obliga a ello.

En cualquiera de ellos, consideremos siempre el valor de la preservación tanto en el evento de desastre futuro como si esto nunca llega a suceder. En el primer caso, tendremos la posibilidad de recuperar el bien original o su copia digital. En el segundo, el simple hecho de tener la información digital accesible y documentada es ya, en sí mismo de gran utilidad.

## 5. PASOS Y FASES DE LA PRESERVACIÓN DIGITAL

El primer paso en preservación digital es, sin ninguna duda, la selección del material que preservaremos y el análisis de sus usos futuros (público objetivo que lo utilizará, motivos y casos de uso). Muchas veces este paso debería ser previo a la propia digitalización en caso de que se trate de artefactos analógicos.

A partir de ahí comienzan las tareas de la preservación en sí, comenzando por la definición de la estructura, metadatos y formatos que utilizaremos y sobre los cuales homogeneizaremos la información. Posteriormente debemos almacenarla de forma ordenada. Después aplicaremos los conceptos de preservación por transformación o de emulación para que ésta sea siempre accesible con la tecnología de la época. Por último, un sistema de preservación debe prever la extracción y reproducción de la información de forma segura.

El modelo de preservación digital más aceptado en la actualidad es el propuesto en OAIS (Open Archival Information System) que tiene su origen en la NASA y es hoy reconocido como un estándar internacional en la norma ISO 14721.

Este modelo se puede sintetizar básicamente en tres etapas:

- 1. Carga (Ingest):** conversión de la información no estructurada en objetos digitales conservables (ODC en adelante), incluyendo los procesos de selección y valoración, validación de la sintaxis propuesta en el plan, integridad, ausencia de elementos dañinos o erróneos, etc.

En *Libsafe* estas tareas están relacionadas con las fases de pre-proceso y de ingestión. También puede ser necesario apoyarse en otras herramientas de gestión de flujos de trabajo para las fases previas (*Libsafe* se apoya en *Libflow*).

- 2. Conservación (archival storage):** consiste en el almacenamiento de los ODC en un sistema seguro, que garantice que no se podrán realizar cambios inadvertidos en los objetos sin que estos sean detectados (en *Libsafe* esto se realiza en el módulo *Archiver*). Adicionalmente incluye las tareas relacionadas con la verificación de que los objetos están “sanos” (de lo cual se ocupa el módulo *Auditor*) y estudia las diferentes estrategias para

su conservación (*Libsafe* implementa la preservación por transformación y se apoya en el módulo *Guardian*).

**3. Recuperación:** al ser almacenados como objetos de un repositorio seguro, los objetos no son accesibles por el usuario directamente, por lo que existen principalmente dos fases en el proceso (OAIS identifica tres diferentes, pero en *Libsafe* sólo se aplican dos, considerando que el formato de almacenamiento es el mismo que el de extracción):

- *Localización del documento:* permite su localización, en base a sus metadatos de origen o los metadatos generados por el propio sistema.
- *Recuperación:* el sistema pondrá a disposición del usuario los datos solicitados, en un área separada de la zona de almacenamiento, de forma que los usuarios puedan acceder, pero no modificar, la información que han solicitado.

## 6. LIBSAFE: HERRAMIENTA DE PRESERVACIÓN DIGITAL A LARGO PLAZO

*Libnova* ha aplicado y extendido el estándar OAIS, adaptándolo a la realidad de archivos, bibliotecas y otros sectores verticales. Hemos desarrollado *Libsafe* para permitir la aplicación de OAIS de una forma eficiente y sencilla.

*Libsafe* es la herramienta perfecta para documentar, diseñar y aplicar un proceso de preservación digital, organizando y protegiendo la información para que esté disponible en el futuro.

Este sistema no sólo nos permite definir y documentar un plan de preservación digital, sino que verifica que la información que se carga es conforme a lo que hemos indicado que debe ser, la limpia y ordena, gestiona los metadatos de los objetos, sus distintas versiones e incluso realiza auditorías automáticas y cíclicas para asegurarse de que todo funciona como debe y de que toda la información preservada es accesible en todo momento.

### *Beneficios de Libsafe*

- **Sencillo de aplicar e implantar:** al ser un producto paquetizado permite la puesta en marcha y ejecución de un plan de preservación digital de una forma sencilla, sin que sea necesario disponer de grandes conocimientos sobre la norma OAIS ni sobre sistemas informáticos.
- **Única herramienta:** organiza todo el proceso de forma que todos los departamentos involucrados quedan vinculados mediante un único sistema de software.
- **Calidad obligatoria:** debido a que *Libsafe* no permite la preservación de elementos que incumplan la política de preservación definida, obliga a todos los agentes que participan en el proceso a que toda la información se introduzca de manera correcta en el sistema.

- **Protección ante obsolescencia:** analiza la tipología de la información (formatos, subformatos, etc.) y muestra un cuadro-resumen del estado de su colección en el tiempo. De esta forma permite anticiparse a los problemas de obsolescencia que una colección pueda sufrir y dispondrá de una sólida base desde la que construir el proceso de migración.
- **Almacenamiento orientado a objetos digitales o entidades intelectuales:** no permite la preservación de ficheros independientes, sino que nos obliga a preservar objetos completos (ficheros, estructura de ficheros y metadatos descriptivos y administrativos). Esto nos da la posibilidad de construir un catálogo de contenido con la opción de buscar sobre él por cualquiera de sus metadatos.
- **Repositorio seguro:** permite establecer una barrera de seguridad entre los usuarios y los archivos de los objetos digitales. Los usuarios nunca podrán acceder a los ficheros que forman parte de los objetos preservados en el sistema. De esta forma se contemplan dos áreas completamente separadas; por un lado el repositorio seguro, y por otro la zona compartida desde la cual el usuario puede acceder a la copia del objeto solicitado. De esta forma, el usuario no podrá en ningún caso modificar el contenido almacenado en el área segura.
- **Versiónado de objetos:** en un sistema de preservación digital nunca se deben borrar objetos que hayan sido preservados. Para permitir la corrección de cambios y evolución de los materiales, *Libsafe* dispone de un sistema de versionado de objetos, mediante el cual podremos ver y recuperar, no sólo la última versión del objeto preservado, sino todas sus versiones anteriores.
- **Notificación de problemas proactiva:** todo es auditado y verificado automáticamente por el sistema. Realiza auditorías automáticas a los objetos del sistema, detectando sin la intervención de ningún operador la modificación o daño de los datos.
- **Inventario permanente:** permite inventariar toda la colección, incluyendo los metadatos de los objetos y generar informes sobre el contenido y el estado de todo el material.
- **Seguridad adicional:** genera una capa adicional de protección, evitando que cualquier operador tenga acceso directo e incontrolado al material preservado.
- **Protección ante desastres:** soporta sistemas geo-distribuidos y colaborativos, realizando automáticamente múltiples copias del material en ubicaciones geográficas diferentes.
- **Basado en estándares:** soporta sistemas de almacenamiento de la práctica totalidad de fabricantes, y guarda toda la información en XML, permitiendo por tanto la migración a diferentes sistemas de almacenamiento de una forma sencilla, garantizando *Libsafe* que todo el proceso se ha realizado correctamente.

## 7. LIBDATA: ALMACENAMIENTO DISEÑADO PARA PRESERVACIÓN DIGITAL

Las necesidades de almacenamiento para preservación son muy diferentes de las que estamos acostumbrados en los entornos de sistemas de información en producción habituales:

- La cantidad de información almacenada en preservación es inmensa en comparación de la que se maneja en producción “en vivo”. En algunos casos, tan sólo el primer proyecto de preservación ya supera la cantidad de información viva total manejada.
- Los sistemas de almacenamiento actuales ponen foco principalmente en la velocidad de transferencia y la accesibilidad permanente, parámetros que son poco importantes para la preservación a largo plazo y encarecen muchísimo el precio.
- Estos sistemas están pensados para modificaciones frecuentes en los datos (información viva) y lectura constante, mientras que los datos preservados no cambian y se acceden muy esporádicamente (la mayor parte de las veces, para auditarlos).
- Un almacenamiento tradicional suele ir acompañado de costosas aplicaciones y procesos de backup, que para preservación no tienen utilidad ya que son difícilmente auditables por lo que la copia preservada es, por concepto, el destino último de los datos.



Teniendo en cuenta estos requisitos Libnova ha desarrollado *Libdata*, un almacenamiento masivo de largo plazo pensado para la preservación de objetos a los que se accede poco frecuentemente pero que necesitamos estar seguros de su disponibilidad en cualquier momento del futuro. Gracias a este diseño se reduce el coste de compra, de operación y de consumos eléctrico y de refrigeración.

*Libdata* se integra con *Libsafe*, que almacena, gestiona, audita y extrae los objetos preservados en estas cabinas. Juntos forman una plataforma sencilla de manejar, flexible y potente en la que implantar el plan de preservación de su organización o crear un repositorio institucional centralizado.

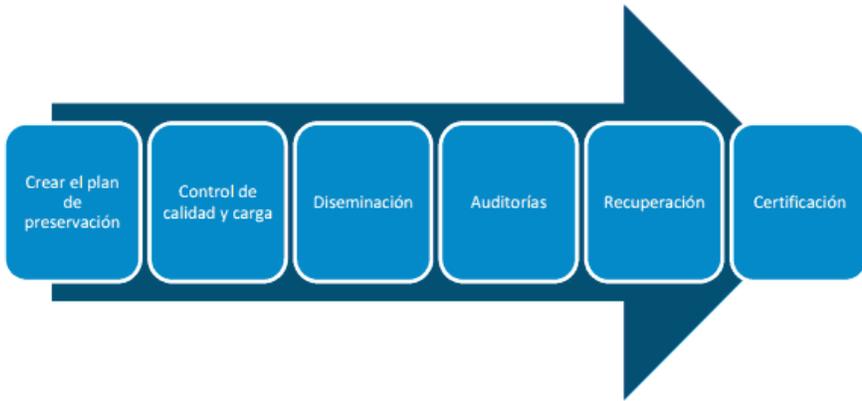
Además, en función del proyecto concreto, *Libnova* puede ofrecer esta plataforma de preservación en modo de pago por uso.

## 8. ORGANIZANDO UN PLAN DE PRESERVACIÓN CON LIBSAFE Y LIBDATA

Desde Libnova colaboramos en la definición y ejecución de planes y proyectos de preservación digital. En estos casos, proponemos contemplar, como mínimo, las siguientes fases:

- **Fase 1 - Estudio del material y elaboración del plan de preservación.** *Libsafe* dispone de un sistema de documentación interno que le permitirá redactar el plan, ofreciéndole además una guía para su elaboración, indicándole todos los puntos a los que debe prestar atención. Todo ello con todas las características que pudiera esperar: con control de versionado y cambios, gestión de permisos, etc.  
Una vez creado este plan de preservación, todo el material de la organización deberá adecuarse a lo propuesto y definido en él.
- **Fase 2 – Verificación y validación.** *Libsafe* realizará una verificación del material a preservar para asegurarse de que cumple con lo especificado en el plan de preservación. El avanzado motor de validación de *Libsafe* puede realizar análisis sobre los metadatos, la estructura de carpetas, realizar validaciones formales de estructura de los archivos (detectando ficheros dañados), analizar su tipología e incluso si la imagen presenta potenciales problemas como por ejemplo estar borrosa (usando filtros de Laplace o transformada de Fourier).
- **Fase 3 – Ingestión y replicación.** En el caso de que el material supere las comprobaciones de la fase anterior, se cargará en el sistema y se realizarán múltiples copias del mismo en los dispositivos de almacenamiento que haya definido. *Libsafe* comprobará de forma totalmente automática y usando funciones de base criptográfica para verificar las copias realizadas, informándole de cualquier anomalía.
- **Fase 4 – Auditorías.** De forma periódica, con la frecuencia y proceso que defina el plan de preservación, la integridad de los objetos preservados debe ser verificada. *Libsafese* ocupa de todo el proceso de forma automática, corrigiendo potenciales errores, realizando copias de los datos en riesgo y avisando al responsable de preservación por correo electrónico en caso necesario. Cuando se integra con la unidad de almacenamiento de datos *Libdata* es, además, capaz de verificar el estado de los discos duros en los que se almacena el objeto digital.
- **Fase 5 – Recuperación.** En caso necesario los objetos pueden ser recuperados utilizando consultas METS. En esta fase es fundamental garantizar la seguridad; *Libsafe* realiza una copia temporal del objeto extraído y la deposita en un almacenamiento intermedio, de forma que el objeto preservado nunca es accesible y está libre de errores de manipulación.
- **Certificación.** No se trata de una fase, ya que se puede realizar en cualquier momento del proceso, y es opción. Desde *Libnova* recomendamos certificar el proceso seguido por una tercera organización distinta de quien lo

realizó. En caso de realizarse, *Libsafe* permite documentar también esta tarea.



## 9. CONCLUSIONES

Desde *Libnova* hemos trabajado para crear un sistema de preservación digital potente, flexible y sencillo de usar que permita a las organizaciones abordar la tarea de la preservación digital de grandes volúmenes de información desde un nuevo punto de vista.