

MÓDULO 7: Gestión y preservación de documentos digitales

SECCIÓN 6
MEDIOS Y
ESTRATEGIAS DE
ALMACENAMIENTO



**ARCHIVO NACIONAL
COSTA RICA**



**UNIVERSIDAD DE
COSTA RICA**

MÓDULO 7

Gestión y preservación de documentos digitales

SECCIÓN 6

Medios y estrategias de almacenamiento

Adaptación del Archivo Nacional de Costa Rica

Versión 1, 2024

Este curso fue traducido y adaptado por la Dirección General del Archivo Nacional de Costa Rica en colaboración con la Sección de Archivística de la Universidad de Costa Rica a partir del material original del año 2011 de la Asociación Internacional de Archivos Francófonos disponible en línea en el Portal Internacional Archivístico Francófono. Se aclara que pueden existir variaciones respecto al contenido original. Para acceder al material en francés, visite <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques>.



Contenido

Capítulo 1. Objetivo de la sección	5
Capítulo 2. Soportes utilizados para la preservación digital	6
2.1. Grabaciones digitales y degradación con el paso del tiempo	6
2.2. Tecnologías existentes	8
2.2.1. Magnético	8
2.2.2. Óptico	8
2.2.3. Flash.....	8
2.2.4. De estado sólido	9
2.2.5. Almacenamiento en nube	9
2.3. Verificación de archivos con suma de comprobación.....	11
2.4. Obsolescencia tecnológica	12
2.5. Condiciones óptimas para la vida útil de dispositivos en un centro de datos.....	14
2.6. Estrategias de preservación digital a largo plazo.....	15
2.6.1. Selección de formatos abiertos y sostenibles en el tiempo	15
2.6.2. Migración y actualización periódica de los formatos	15
2.6.3. Uso de metadatos oportunos y consistentes.....	15
2.6.4. Almacenamiento seguro y redundante	15

Capítulo 3. Estrategias de almacenamiento	16
3.1. Entidad “almacenamiento” del modelo OAIS	17
3.2. Abstracción de la plataforma de hardware	18
3.3. Jerarquía de almacenamiento	19
3.4. Características de los servicios de almacenamiento.....	21
Capítulo 4. Políticas de almacenamiento	22
4.1. Aplicación interna	22
4.2. Mutualización	22
4.3. Externalización (terceros archivadores)	23
Capítulo 5. Casos de estudio	23
5.1. Proyecto de archivos de la UNESCO	23
5.2. Proyecto de digitalización del patrimonio cultural de Google	23
5.3. Europea	24
Bibliografía	24

Capítulo 1. Objetivo de la sección

En definitiva, será siempre el archivista quien tendrá la responsabilidad de la conservación de la información. Esta sección 6 está dedicada a los soportes y a las estrategias de almacenamiento, debe permitir al archivista una visión clara de los posibles riesgos y soluciones. En la mayoría de los casos, la aplicación de las soluciones estará a cargo de los informáticos, pero el archivista deberá ser capaz de sopesar las ventajas y los inconvenientes de las posibles soluciones y tomar decisiones con pleno conocimiento de causa.

La sección se presenta desde el punto de vista de la preservación, de las secuencias de bits y de la integridad de estas secuencias durante el tiempo.

Las soluciones o las estrategias de almacenamiento que se adopten se apoyarán evidentemente sobre las tecnologías disponibles en este campo.

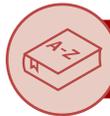
Sin embargo, no hay que creer que se trata de una cuestión puramente técnica. Las decisiones se basarán en un análisis de las necesidades y limitaciones presentes y futuras del Archivo y de su contexto.

Nos encontramos en un contexto de la revolución digital que nos permite acceder a medios de almacenamiento más amplios y asequibles en comparación con décadas pasadas, permitiéndonos el desarrollo de las actividades de preservación de datos de una manera más efectiva.

Uno se pregunta:

- ¿Dónde resguardar los datos?, ¿En cuál repositorio?, ¿Qué medios utilizar? Así mismo se debe analizar también como organizarlos, como preservarlos en el tiempo, garantizando que no se perderán y asegurando también su integridad.
- ¿Cómo sobrevivir a los incidentes materiales y al envejecimiento de los soportes?
- ¿Cómo tener garantizar su integridad al pasar de un medio a otro?
- ¿Cómo asegurar que los datos no sean modificados ni leídos de forma no autorizada?

Las infraestructuras de almacenamiento deberán tener en cuenta una serie de requisitos relativos a las prestaciones, los tiempos de acceso a los documentos, la seguridad y, al mismo tiempo, respetar las limitaciones de costes.



GLOSARIO

Archivo: Organización responsable de conservar la información para que una comunidad de usuarios objetivo pueda acceder a ella y su utilización (glosario del modelo de referencia OAIS).

Huella (huella digital o condensado o hash): Resultado de una función hash aplicada a una Cadena de caracteres de cualquier longitud para reducirla a una longitud fija representativa de esta cadena de caracteres. La huella es uno de los elementos que permiten verificar la integridad de un documento, de un flujo, de un lote, de una transmisión, (comparación de huellas dactilares).

Petabyte: Unidad de medida de almacenamiento cuyo símbolo se representa por el PB, equivale a 1000 terabytes.

Zettabyte: Unidad de medida de almacenamiento cuyo símbolo se representa por el ZB, equivale a 1 millón de petabytes y a 1000 millones de terabytes.

Capítulo 2. Soportes utilizados para la preservación digital

Los soportes de preservación digital son los medios o dispositivos utilizados para almacenar y mantener los datos digitales a largo plazo, entre estos podemos encontrar los discos ópticos, cintas magnéticas, discos duros magnéticos (HDD), discos duros de estado sólido (SSD) o el almacenamiento en nube.

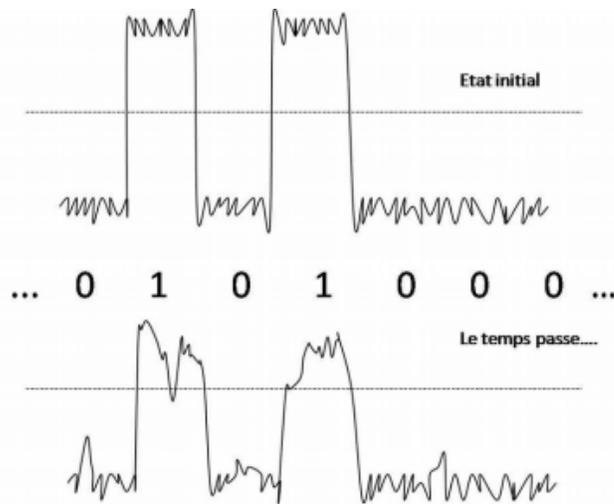
Cada uno de estos soportes posee características únicas relacionadas a su longevidad, capacidad de almacenamiento, escalabilidad, velocidad de acceso y recuperación de los datos, se debe de tener en cuenta además su costo y sostenibilidad a través del tiempo.

La elección del soporte a utilizar debe de ir acompañada de una buena práctica de gestión y mantenimiento para garantizar la integridad y accesibilidad a largo plazo de los datos resguardados sin ningún impedimento de por medio que arriesgue la recuperación de la información.

2.1. Grabaciones digitales y degradación con el paso del tiempo

La grabación de información digital en un soporte con el fin de memorizar o conservar permanentemente esta información da como resultado la grabación en este soporte de secuencias de 0 y 1. Esta inscripción se realiza bajo una forma codificada, la codificación puede depender de la tecnología utilizada.

La realidad física implica la inscripción de valores discretos en el soporte, esta inscripción generalmente se degradará de manera progresiva con el tiempo.



Estado inicial / Tiempo transcurrido

Degradación de señal registrada con el tiempo. Elaborado por PIAF

En la era digital, el almacenamiento y la preservación de los datos han experimentado una evolución significativa. Las grabaciones digitales han reemplazado en su mayoría a su contraparte analógica ofreciendo entre muchos aspectos de mejora, mayor flexibilidad, calidad y respaldo a través del tiempo. A pesar de sus ventajas, las grabaciones digitales están lejos de ser perfectas, como parte de sus desafíos se enfrentan las condiciones del paso del tiempo, es por esto por lo que tenemos que tomar en cuenta una serie de factores que pueden influir en esta degradación:

- **Pérdida de metadatos:** Los metadatos básicos de los archivos digitales, tales como la fecha de creación, ubicación, hora, entre otros, pueden perderse dificultando la identificación y sus de datos de referencia.
- **Compresión y codificación:** La compresión y codificación pueden afectar la calidad original de los objetos digitales, según los algoritmos de compresión utilizados es posible obtener un resultado de degradación y pérdida de calidad mayor al esperado para su conservación a largo plazo.
- **Errores de almacenamiento:** Las grabaciones digitales pueden verse corrompidas por errores de almacenamiento en sectores dañados en discos duros, los errores de escritura y lectura en los diversos medios de almacenamiento o problemas de conexión durante la transferencia en nube son otros ejemplos.
- **Degradación física del soporte:** Si bien los medios digitales no están sujetos a la degradación física de los medios analógicos, tales como las cintas magnéticas, aún pueden deteriorarse con el tiempo, las memorias USB, por ejemplo, pueden sufrir deterioro al exponerse a la luz ultravioleta, la humedad o cambios de temperatura extremos.
- **Obsolescencia tecnológica:** Los formatos y dispositivos de reproducción han reflejado la rapidez en la que estos pueden volverse obsoletos en un corto período de tiempo, por ejemplo, medios físicos como los CDs o DVDs pueden deteriorarse físicamente y están sujetos a la necesidad de un reproductor en buen estado, dispositivo que recientemente se ha eliminado de los equipos personales actuales.

2.2. Tecnologías existentes

Existen varios tipos de tecnologías para el soporte del almacenamiento de datos, con sus distintas cualidades, entre estos se encuentran:

2.2.1. Magnético

- **Cintas Magnéticas:** Las tecnologías magnéticas son las más antiguas, sí bien han surgido diversas opciones para la grabación de datos, esta continúa vigente en la actualidad, se trata de una cinta recubierta con material magnético que se enrolla dentro de un carrete para almacenar datos de manera secuencial. A partir de los años 1990, se popularizó el uso de las cintas LTO (Linear Tape-Open) el cual en la actualidad se encuentra en su versión número 9 con capacidad de almacenamiento de hasta 45 TB por cartucho de forma comprimida.
- **Discos Duros (HDD):** Desde su existencia en 1960 se ha convertido en una de las principales tecnologías para la grabación de datos, inicialmente teniendo estos un costo elevado y menor capacidad hasta la actualidad donde es el recurso favorito de los centros de datos a nivel mundial, han logrado posicionarse como un recurso de gran fiabilidad y escalabilidad.

Los HDD son conocidos como discos duros tradicionales, estos dispositivos utilizan discos magnéticos giratorios para escribir y leer información de manera persistente. Este dispositivo se destacaba por tener una gran capacidad de almacenamiento la cual puede alcanzar los 20 TB. Esta tecnología cuenta con una amplia longevidad, soportando un gran número de ciclos de escritura y lectura antes de presentar fallas, además, los HDD ofrecen una capacidad de almacenamiento alta a un costo menor, en comparación con otros servicios o tecnologías.

2.2.2. Óptico

- **Digital Versatil Disc (DVD):** Disco óptico el cual hace uso de la tecnología láser para la escritura y lectura de datos en una capa reflectante de plástico, sí bien debido a la creciente popularidad de tecnologías más accesibles y de mayor capacidad, el uso del DVD ha disminuido en gran manera en los últimos años, aunque estos se continúan utilizando en ciertas industrias se trata de una alternativa actualmente en declive.
- **Blue-ray Disc:** Similar al DVD, el Blue-ray utiliza un láser de color azul-violeta para la escritura y lectura de datos en una capa de plástico de alta densidad, permitiendo una mayor capacidad que su antecesor, una unidad estándar puede alcanzar los 50 GB y hasta 100 GB en sus versiones de 4 capas. Desde su nacimiento en 2005, se enfrenta a los grandes retos de ser el sucesor principal de los discos ópticos bajo un esquema de transición hacia los medios extraíbles USB y el almacenamiento en nube, sin embargo, aún posee un nicho de mercado enfocado en la calidad de audio y vídeo.

2.2.3. Flash

- **Memoria USB:** La memoria USB es un tipo de tecnología para grabación de datos no volátil que se puede utilizar en gran variedad de dispositivos electrónicos, se ha destacado enormemente por sus características de rápido acceso, durabilidad y bajo consumo de energía, teniendo capacidad para transmitir datos incluso entre teléfonos celulares inteligentes. Su popularidad como uso personal ha logrado colocarlo como la tecnología con mayor uso en los últimos años, esto debido a sus ventajas frente a los medios ópticos, frente a agentes como el polvo o rasguños, así como su amplia capacidad que alcanza los 4TB en la actualidad.

2.2.4. De estado sólido

- **Unidades de Estado Sólido (SSD):** Las unidades SSD (Solid State Drive) son dispositivos de almacenamiento que utilizan un tipo de chips de memoria flash NAND, utilizado para la escritura y lectura de datos de manera persistente, su principal diferencia en comparación a los discos duros tradicionales es que estos no utilizan discos magnéticos giratorios, dando como resultado tiempos de acceso mucho más rápidos, así como tasas de transferencia más altas. Algunos se refieren erróneamente a estos dispositivos como “discos”, sin embargo, a diferencia de sus predecesores, sus datos ya no se almacenan sobre superficies cilíndricas.

Entre sus ventajas se destaca:

- **Rendimiento:** Mayor capacidad de acceso y transferencia de datos.
- **Eficiencia energética:** Poseen un menor consumo que los discos duros tradicionales, aspecto que beneficia a los dispositivos portátiles tales como laptops o tablets.
- **Fiabilidad:** Al no poseer partes móviles dentro de su estructura, los SSD son menos susceptibles a daños por golpes y vibraciones.
- **Tamaño Compacto:** A raíz de la falta de su diseño sin partes móviles, los SS son más compactos y ligeros que los discos duros tradicionales.

A pesar de que los SSD poseen numerosas ventajas sobre los discos duros tradicionales, estos poseen algunos retos tales como las limitaciones en su capacidad de almacenamiento, el cual suele ser inferior a los discos duros tradicionales, así como su diferencia en costo, otro aspecto a tomar en cuenta es la degradación de estos dispositivos los cuales tienen un límite de escrituras y lecturas.

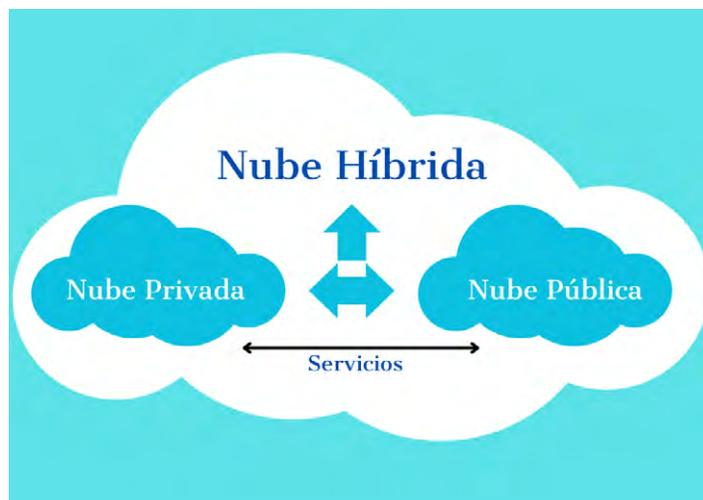
2.2.5. Almacenamiento en nube

- **Almacenamiento basado en la nube:** Se trata de una tecnología de grabación de datos que ha ganado gran popularidad en los últimos años. Los datos se almacenan en servidores remotos en internet desde donde se puede obtener acceso utilizando distintos dispositivos inteligentes, evitando de esta manera guardar los datos en dispositivos locales tales como unidades de disco duro o unidades flash.

Entre las principales particularidades del servicio de almacenamiento en la nube se encuentra su accesibilidad, permitiendo a los usuarios administrar sus datos desde cualquier lugar, mientras se posea conexión a internet. La mayoría de estos servicios permite respaldar de manera automática la carpeta asociada a la nube, además de realizar copias de seguridad y un control de versiones, permitiendo recuperar archivos eliminados de manera errónea y auditar las actividades realizadas en este espacio.

Existen tres tipos principales de almacenamiento en la nube:

1. **Almacenamiento en la nube privada:** Cuando se almacenan datos en un centro de datos privado, utilizada por una única organización.
2. **Almacenamiento en la nube pública:** Cuando se almacenan datos en un centro de datos público compartido entre varias organizaciones, este a su vez administrado por un proveedor representante de la marca.
3. **Almacenamiento en la nube híbrida:** Cuando se realiza una combinación entre el almacenamiento en nubes públicas y privadas, normalmente se utilizan servicios de manera integrada.



Representación de tipos de almacenamiento en nube. Elaboración propia

La revolución digital ha impulsado el uso de esta tecnología en los últimos años, según un informe (CyberSecurity-Ventures, 2020) para el año 2025 se alcanzará un almacenamiento total de 100 zettabytes de datos en la nube. Esta investigación también evidencia que el 50% de los individuos hacían uso del almacenamiento en la nube, ya sea como uso personal o laboral, es por esto que organizaciones han fijado su atención sobre el uso de almacenamiento en la nube como parte de su infraestructura.

Este servicio dependerá de una conexión estable a internet, ya que, aunque existen servicios de almacenamiento en la nube que realizan una sincronización automática de los cambios recientes, la cantidad de cambios pendientes por sincronizar puede aumentar el tiempo de duración de este proceso, el cual puede limitarse según el ancho de banda al que se conecte el equipo.

Estas tecnologías de almacenamiento de datos poseen sus propias características, su elección dependerá del tipo de dato que requiera almacenar entre varios factores tales como:

- **Fiabilidad:** Capacidad de la tecnología de almacenamiento para evitar la pérdida o corrupción de los datos.
- **Durabilidad:** Capacidad de los medios para conservar el almacenamiento de los datos durante un prolongado periodo de tiempo.
- **Seguridad:** La seguridad de los datos es crucial, se debe de seleccionar una tecnología de grabación que ofrezca una protección adecuada contra filtraciones de datos, robos o accesos no autorizados.
- **Costo:** Las diferentes tecnologías de almacenamiento poseen distintos costos, es importante seleccionar una tecnología que se adapta al estado actual de nuestra organización mantenimiento la mejor relación calidad-precio.
- **Compatibilidad:** El tipo de tecnología a utilizar puede ser incompatible con ciertos formatos u otras tecnologías. Por ejemplo, la migración de datos en cintas magnéticas a almacenamiento en nube podría complicarse por su incompatibilidad y tiempo requerido para transferir los datos.
- **Escalabilidad:** Las necesidades organizacionales pueden variar con el tiempo, es importante seleccionar una tecnología de almacenamiento que sea escalable y que sea adaptable a las posibles crecientes necesidades de almacenamiento de los datos.

- **Accesibilidad** La accesibilidad es un factor crítico por considerar, la tecnología de almacenamiento a utilizar debe de permitir una fácil recuperación de datos cuando estos sean necesarios, sin comprometer su integridad.



COMPLEMENTO

El medio digital alemán DW, público una noticia en febrero de 2024 titulado: Científicos chinos desarrollan disco óptico de 1,6 petabytes, haciendo referencia al posible resurgir de esta tecnología de grabación de datos.

<https://www.dw.com/es/cient%C3%ADficos-chinos-desarrollan-disco-%C3%B3ptico-de-16-petabytes/a-68338903>

La página web de origen español Hardzone, publicó en julio de 2023 un reportaje sobre las memorias NAND, detalles acerca de su funcionamiento y características principales.

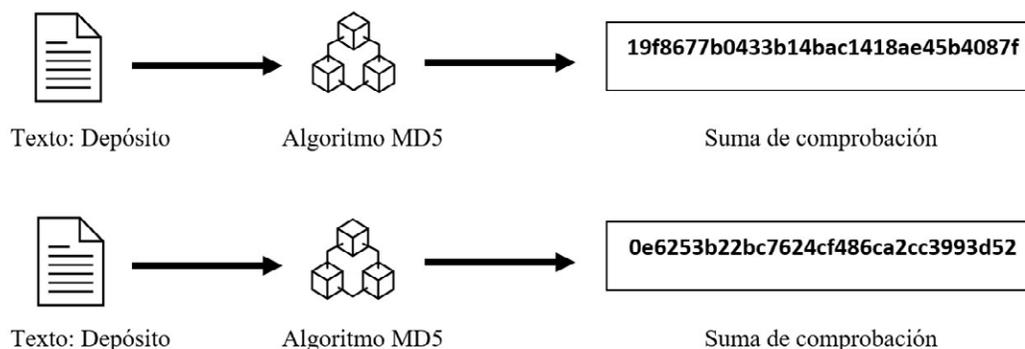
<https://hardzone.es/reportajes/que-es/memoria-nand-flash/>

2.3. Verificación de archivos con suma de comprobación

La suma de comprobación es una cadena de caracteres utilizado para referenciar de manera única a un archivo. Este procedimiento se utiliza para verificar si un archivo conserva su integridad de datos durante su transmisión o si este ha sido alterado. Este valor se genera de tal manera que el mínimo cambio en los datos del archivo resultará en una suma de valor totalmente distinta.

Para realizar una suma de comprobación, se debe procesar el archivo utilizando un algoritmo adecuado, dando como resultado una cadena alfanumérica única comúnmente llamado “checksum” o suma de comprobación.

Los algoritmos de suma de comprobación más comunes son SHA-1, SHA-256, MD5, entre otros, pueden ser utilizadas en archivos por medio de aplicaciones en línea, así como en versiones de escritorio.



Ejemplo de suma de comprobación con resultados diferentes. Elaboración propia a partir de PIAF

Una vez obtenida la suma de comprobación del archivo original se debe realizar un segundo proceso con el archivo a comparar, contrastando ambas cadenas de caracteres es posible determinar si el archivo copia ha sufrido alguna alteración y si así se desea, es posible investigar la causa de esta discrepancia.

Las sumas de comprobación se utilizan en escenarios como:

- Verificación de la integridad de archivos en un sistema de almacenamiento y sus copias de seguridad.
- Verificación de la integridad de archivos durante la transferencia de datos (envió por correo electrónico o archivos en la nube)
- Verificación de la integridad de archivos descargados de internet y distribución de software.
- Auditorías de seguridad a sistemas o sistema de almacenamiento.

En resumen, la suma de comprobación posee una amplia gama de aplicaciones para la verificación de la integridad y autenticidad de los archivos. Actualmente se trata de una práctica común para garantizar la seguridad y fiabilidad de los archivos en un repositorio digital.

2.4. Obsolescencia tecnológica

La obsolescencia de los soportes digitales es un problema de gran índole en la industria de la tecnología, información y comunicación. Se refiere al riesgo al que están expuestos los datos almacenados en un formato particular frente a la posibilidad de volverse inaccesibles debido a tecnología actual asociada con ese formato.

Las causas de este fenómeno pueden ser variadas, entre estas podemos destacar:

- **Cambios en los estándares de la industria:** Los estándares de almacenamiento y comunicación como los tipos de dispositivo, interfaces de conexión y protocolos de transferencia, pueden variar con el tiempo. Si un archivo depende de un estándar que se vuelve obsoleto, los dispositivos y software modernos pueden no ser compatibles con ese archivo.
- **Formatos de archivo obsoletos:** A medida que evoluciona la tecnología, es posible que se establezcan nuevos formatos de archivo con mayor calidad, compresión o eficiencia, provocando que los formatos de archivo más antiguos se vuelvan obsoletos o con dificultades para la manipulación o lectura de los datos almacenados.
- **Degradación física de los medios:** Todos los medios de almacenamiento tienen una vida útil limitada y pueden deteriorarse al exponerse a ciertos factores como la exposición al calor, humedad o luz ultravioleta. Esto puede resultar en una pérdida de los datos y la obsolescencia de los medios.
- **Descontinuación de productos y servicios:** En ocasiones los fabricantes dejan de producir dispositivos de almacenamiento y software que es compatible con ciertos medios de formatos digitales. Esto puede hacer que los formatos se vuelvan obsoletos y difícil de mantener a futuro.



EJEMPLO

- El formato digital MPEG-2 (Moving Picture Experts Group -2) fue una vez el estándar de compresión de video digital y audio, desarrollado en 1995 como una mejora de su predecesor, MPEG-1, fue ampliamente utilizado para la televisión digital, DVDs y la transmisión en directo por medio de internet. Aunque este no se ha vuelto del todo obsoleto, su uso ha disminuido por la creación de formatos más modernos como el H.264 (AVC) y H.265 (HEVC).
- El formato Windows Media Audio (WMA), desarrollado por Windows, es otro claro ejemplo de la obsolescencia en formatos digitales. Aunque fue ampliamente utilizado décadas atrás, fue reemplazado por formatos más populares y universales como el MP3 y el AAC.

Con el fin de mitigar los riesgos asociados con la obsolescencia en los soportes digitales, es importante tomar medidas proactivas y estar a la vanguardia sobre la información actualizada en esta materia. Realizar tareas como la migración a formatos o medios modernos y gestionar copias de seguridad de estos archivos es parte de las alternativas recomendadas, además de la adopción de formatos estándar de preservación a largo plazo o de software libre, tales como el PDF/A, TIFF o MKV.



COMPLEMENTO

Ejemplos de formatos digitales de preservación a largo plazo

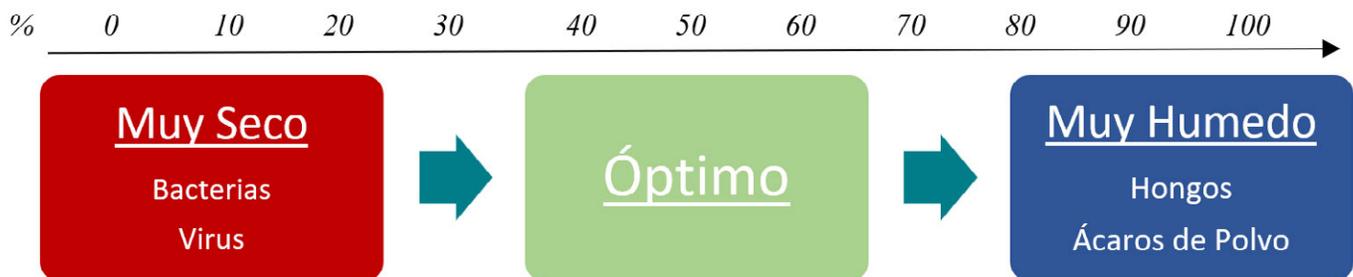
Formato	Descripción	Ventajas	Limitantes
PDF/A	Formato PDF diseñado para la preservación a largo plazo de documentos electrónicos.	<ul style="list-style-type: none"> • Amplia Compatibilidad. • Protección contra inserción de URL, código Javascript, multimedia, entre otros. 	<ul style="list-style-type: none"> • Dependencia de fuentes incrustadas en el documento. • Limitaciones con características dinámicas presentes en algunos PDF estándar.
TIFF	Formato de imagen sin pérdidas, diseñado para el almacenamiento de gran calidad.	<ul style="list-style-type: none"> • Conserva la calidad de la imagen original. • Soporta metadatos descriptivos. 	<ul style="list-style-type: none"> • Problemas de almacenamiento derivado del gran tamaño de los archivos.
FLAC	Formato de compresión de audio sin pérdidas para la preservación de la calidad original.	<ul style="list-style-type: none"> • Posee gran compatibilidad con los distintos reproductores de audio. • Conserva la calidad del sonido original. 	<ul style="list-style-type: none"> • Es posible que los archivos FLAC sean más pesados que otros formatos de compresión.
MKV	Formato de compresión utilizado para almacenar audio, video y subtítulos.	<ul style="list-style-type: none"> • Soporta metadatos descriptivos. 	<ul style="list-style-type: none"> • Posible incompatibilidad con algunos reproductores de medios.

2.5. Condiciones óptimas para la vida útil de dispositivos en un centro de datos

La duración de los recursos físicos presentes dentro de un centro de datos está estrechamente relacionada con las condiciones y medidas de protección que estos posean. Un entorno deficiente puede disminuir rápidamente el tiempo de vida útil de los dispositivos, causando daños catastróficos e irreversibles si no se detecta y se corrige en un tiempo preciso.

A continuación, se brindan algunas recomendaciones que procuran alargar la vida útil de los dispositivos en los centros de datos:

- **Temperatura:** Mantener una temperatura estable entre 18-27°C, se debe de evitar el sobrecalentamiento, así como el enfriamiento excesivo, ya que estas condiciones pueden causar daños en los dispositivos.
- **Control de polvo y otras partículas:** Se debe minimizar la cantidad de polvo y partículas en el aire, con el fin de prevenir que estos se acumulen en los disipadores y otros componentes de los dispositivos, la instalación de filtros de aire es una solución eficaz para este punto.
- **Alimentación eléctrica estable:** Los dispositivos de los centros de datos requieren una alimentación eléctrica estable y de gran fidelidad para funcionar de manera correcta. La instalación de reguladores de voltaje, suministros eléctricos redundantes y UPS de doble conversión, son medidas eficientes para proteger los equipos ante cortes eléctricos y picos de voltaje.
- **Humedad relativa:** La humedad relativa previene la acumulación de estática y corrosión en los componentes eléctricos de los dispositivos, así como la generación de descargas electrostáticas causadas por la escasez de humedad. Se recomienda mantener una humedad relativa entre el 40% y 60% dentro del entorno de centro de datos, la instalación de un dispositivo deshumidificador es de utilidad para controlar este parámetro.



Estados según el porcentaje de humedad relativa en un centro de datos. Elaboración propia

- **Monitoreo y mantenimiento:** Se debe llevar a cabo un monitoreo continuo tanto de las condiciones ambientales del área, así como el rendimiento y alertas que los equipos o sus componentes posean.

Atender estas recomendaciones beneficia el tiempo de vida útil de los dispositivos en los centros de datos, es importante anticipar y atender los riesgos que afecten la funcionalidad del recinto, así como emplear buenas prácticas como la adquisición de repuestos o consumibles de reserva para aplicar las correcciones presentadas en el menor tiempo posible.

2.6. Estrategias de preservación digital a largo plazo

La preservación de los recursos digitales es un proceso de suma importancia, no es suficiente tener un plan de creación y administración de respaldos de la información, existe la necesidad de garantizar la accesibilidad e integridad de estos medios a través del tiempo.

Algunas estrategias recomendadas para lograr este objetivo son:

2.6.1. Selección de formatos abiertos y sostenibles en el tiempo

Dar prioridad al uso del formato de archivos abiertos, que sean reconocidos por distintas entidades como ampliamente compatibles y con poca posibilidad de volverse obsoletos a corto plazo. Pueden considerarse como parte de estos formatos los observados en la tabla: Ejemplos de formatos digitales de preservación a largo plazo, del capítulo 2.4.

En cuanto a los estándares relacionados con este punto, la norma ISO 19005 – Gestión de documentos - Formato de fichero de documento electrónico para la conservación a largo plazo, define las directrices para la creación y el uso de formatos de archivos electrónicos que son óptimos para la preservación de documentación a largo plazo. Esta norma específica en sus diferentes secciones, el uso del formato PDF/A y las características de sus variantes.

2.6.2. Migración y actualización periódica de los formatos

Se debe de implementar procesos de migración y actualización de formatos para asegurar que los recursos digitales se mantengan accesibles a medida que se da la evolución de nuevos estándares y tecnologías.

2.6.3. Uso de metadatos oportunos y consistentes

Añadir al plan de trabajo, la aplicación y mantenimiento de metadatos descriptivos y técnicos para los recursos digitales que se gestionen en la organización. Estos metadatos facilitarán la identificación, búsqueda y el contexto general de los archivos digitales a lo largo del tiempo, contribuyendo con su preservación y lectura.

Como marco de trabajo básico para la implementación de metadatos, el conjunto Dublin Core¹, basado en la norma ISO 15836:2009 - Información y documentación., nos brinda las directrices para la implementación de metadatos descriptivos para recursos digitales, como documentos, imágenes, videos, sitios web, entre otros. Estos elementos proporcionan información básica por medio de etiquetas sobre el curso, como su título, autor, fecha, tipo de contenido, entre otros, facilitando su búsqueda, gestión y recuperación, componentes fundamentales para su preservación a largo plazo. (Ver sección 9 dedicada a los Metadatos)

2.6.4. Almacenamiento seguro y redundante

Se recomienda el uso de sistemas de almacenamiento seguro y redundante para la protección de los recursos digitales, se debe priorizar el aseguramiento de los archivos contra condiciones que los pongan frente a riesgos de corrupción de datos, pérdidas parciales o totales, además de los accesos no autorizados.

1 <https://www.dublincore.org/>

Es posible emplear soluciones físicas para el almacenamiento distribuido, la utilización de discos duros redundantes e incluso servicios de almacenamiento en nube, que brinden las características de seguridad y accesibilidad requeridas para los recursos.

La norma ISO 16363:2017-Sistemas de transferencia de información y datos espaciales - Auditoría y certificación de repositorios digitales de confianza, establece los requerimientos para auditar y certificar los repositorios digitales confiables para la preservación de datos a largo plazo. Esta norma evalúa que se cumplan con los estándares necesarios que garanticen la accesibilidad, autenticidad e integridad de los recursos digitales alojados en un repositorio de largo plazo.

Entre los aspectos que contempla esta norma se encuentran:

- Las definiciones y términos clave relacionados con la preservación de los datos y los repositorios digitales confiables.
- Requisitos generales para la administración de repositorios digitales, incluyendo las políticas y procedimientos.
- Requerimientos para la gestión de datos y metadatos, incluyendo la documentación de los recursos digitales y la preservación de los metadatos técnicos y descriptivos.
- Procedimiento de auditoría y certificación para evaluar concordancia de un repositorio con los requisitos establecidos por la norma.

En conclusión, combinando estas estrategias, es posible desarrollar un enfoque integral basado en normativas actuales, que nos brinde los pasos iniciales para la preservación de los recursos digitales a largo plazo, asegurando de manera proactiva su integridad y accesibilidad de consulta para los usuarios.

Capítulo 3. Estrategias de almacenamiento

Almacenar es sólo una de las funciones del archivo y no debe confundirse. Si el almacenamiento es un elemento esencial del proceso, en ningún caso constituye el único elemento de una estrategia de sostenibilidad, pero sí es la base en que se basa el archivo.

¿Qué se espera de un servicio de almacenamiento dentro de un archivo digital?

Es una infraestructura que es capaz de asegurar la garantía de una restitución de información.

- Sin alteraciones (integridad, autenticidad)
- A largo plazo (migración, obsolescencia)
- En un problema de coste controlado

En primer lugar, repasemos con más detalle lo que nos dice el modelo OAIS sobre la entidad funcional “Almacenamiento”.

3.1. Entidad “almacenamiento” del modelo OAIS

La entidad “almacenamiento” del modelo OAIS asegura las siguientes funcionalidades:

1. Recibir los AIP (paquetes de información de archivo)

Registrar en una zona de almacenamiento adecuada y enviar un certificado de que el almacenamiento se ha realizado.

2. Gestionar la jerarquía de almacenamiento

En función de los requisitos de calidad de los servicios de las entidades de pago y de acceso, en función de la frecuencia si se conoce la utilización prevista de la AIP, se seleccionarán los soportes adecuados que permitan el acceso en línea, en diferido o ligeramente diferido. Deberá comprobarse previamente que los AIP no se han alterado durante la transferencia

Esta función también proporciona estadísticas sobre los medios disponibles y la capacidad de almacenamiento disponible en las diferentes capas, así como sobre la utilización de los AIP.

3. Asegurar la migración de soportes

Esta función ofrece la posibilidad de reproducir los AIP durante el tiempo. Estas son las migraciones tecnológicas. En curso de las operaciones, el contenido informativo y la información perenne, no deben de sufrir modificaciones

Sin embargo, los datos que constituyen la información de empaquetado pueden ser cambiados, siempre que sigan desempeñando la misma función. La estrategia de migración consistirá en elegir un medio de almacenamiento teniendo en cuenta los porcentajes de error reales y esperados que caracterizan a los distintos tipos de soportes, su rendimiento, y su coste de adquisición.

4. Control periódico de la integridad de la información confiada

Esta función garantiza, con una probabilidad estadísticamente aceptable, que ningún AIP ha sido corrompido en una transferencia interna de los datos de la entidad “almacenamiento”. La información de integridad asegura, en cierta medida, que el contenido de información no ha sido modificado a lo largo del tiempo, independientemente de si se ha desplazado o no el AIP de un medio a otro, con independencia de que se acceda o no a dicho AIP.

5. Regenera los datos en caso de siniestro

El «plan de recuperación» proporciona un mecanismo para duplicar el contenido digital del Archivo y almacenar la copia en una instalación geográficamente distante.

6. Proporciona los AIP a la entidad de acceso

Esta función transmite a la Entidad «Acceso» las copias de los AIP almacenados. Esta función recibe una solicitud de AIP especificando los AIP solicitados y los libros sobre el tipo de soporte solicitado o los transfiere a un espacio de almacenamiento provisional.

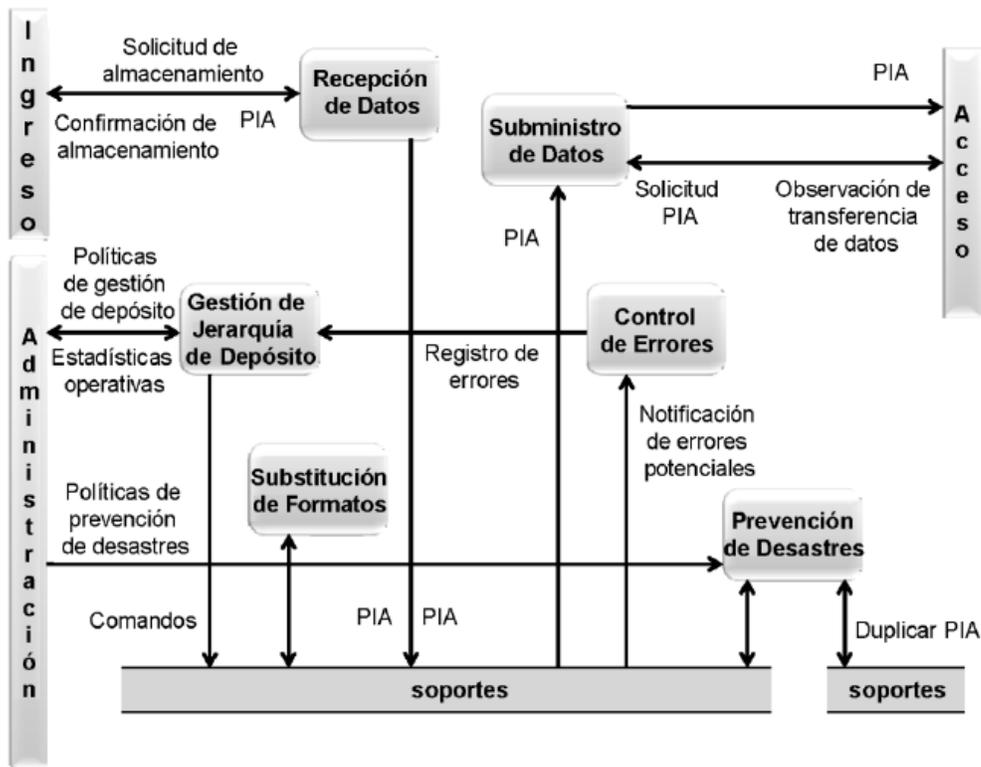


Figura 4-3 – Funciones de la Entidad Funcional Almacenamiento de Archivo

ISO 14721 (2015) p.46. Funciones de la entidad funcional: almacenamiento

Vamos a encontrar todas estas características en la implementación de un servicio de almacenamiento.

3.2. Abstracción de la plataforma de hardware

El almacenamiento adopta la forma de un conjunto de medios organizados en sistema. Un “sistema de almacenamiento” es un conjunto de software y hardware que proporcionan un servicio plenamente definido.

Sabiendo que la plataforma va a evolucionar regularmente ya que está constituida por hardware y software los cuales van a cambiar con el tiempo, se tendrán que gestionar operaciones de vigilancia y renovación de medios de almacenamiento, parece que hay un gran interés en constituir esta plataforma de forma autónoma en comparación con el resto del sistema de archivo digital.

Este enfoque pragmático se basa principalmente en la experiencia de las plataformas existentes.

Con un enfoque de este tipo:

- Las evoluciones de la plataforma de almacenamiento no deben afectar la organización logística del archivo: el productor de los datos y también el archivo gestionan una organización lógica de los datos y la capacidad de transformar estos datos en información inteligible. Esta organización lógica no tiene correspondencia con la organización de la arquitectura de almacenamiento utilizada, organización que puede ser migrada con el tiempo en función de tecnologías y de los servicios disponibles.

- Los cambios de política de la jerarquía de almacenamiento (clases de servicio) y las migraciones de la arquitectura base del repositorio deberán ser transparentes, es decir, no repercutir en las entidades usuarias del servicio de almacenamiento.

Por lo tanto, los usuarios (también llamados clientes) del servicio de almacenamiento, no necesitan saber bajo que arquitectura de almacenamiento se resguardan sus datos. Sin embargo, necesitan disponer de las garantías del servicio que se les brinda bajo los términos de sostenibilidad, integridad y accesibilidad.

Se trata de un enfoque de servicio. Es, por ejemplo, lo que esperamos de un servicio de paquetería, esperamos que nuestro paquete llegue a su destinatario, en buen estado y en el tiempo previsto. No importan los medios de transporte que el proveedor del servicio utilice para transportar el paquete. Este diseño facilita también la puesta en común, es decir, la utilización del servicio de almacenamiento por diferentes grupos, departamentos, proyectos dentro de la organización e incluso entre organismos asociados.



COMPLEMENTO

La NASA utiliza una combinación de sistemas de almacenamiento en nube, sistemas de almacenamiento de alto rendimiento y sistemas de almacenamiento a largo plazo para gestionar sus datos científicos y de investigación de misiones espaciales entre otros medios digitales. Para esto hacen uso del Centro de Distribución de la Tierra (EOSDIS)² por sus siglas en inglés, para almacenar y distribuir esta información.

En el artículo titulado “Earth Science Data in the Cloud: The EOSDIS Cumulus Project” (Blumenfeld, 2017) Se comenta sobre la evolución de este proyecto y la transición de estos servicios hacia la plataforma en nube, Blumenfeld hace mención, además, sobre el estado actual del almacenamiento de los datos recolectados, el cual se estimaba en 22 petabytes (PB) para esta fecha y se esperaba un crecimiento de hasta 247 PB para el 2025, dando una clara referencia de la enorme cantidad de datos procesados y recolectados como resultado de las operaciones de este programa de la NASA.

3.3. Jerarquía de almacenamiento

Teniendo en cuenta que no todos los recursos digitales poseen la misma trascendencia para la organización y que existen algunos más consultados que otros, la jerarquía de almacenamiento nos brinda un enfoque estratégico con el cual es posible realizar una clasificación de los recursos digitales en diferentes niveles de almacenamiento, optimizando la preservación a largo plazo sin comprometer la accesibilidad o seguridad de los archivos resguardados.

Bajo esta jerarquía, los recursos digitales se organizan en diferentes niveles de almacenamiento, desde el más seguro y duradero, hasta el de menor costo y con más susceptibilidad a fallos. Podríamos optar en la parte superior de la jerarquía por medios físicos de almacenamiento de alta durabilidad y calidad, tales como discos duros o cintas magnéticas. En contraste, en la parte inferior de la jerarquía podríamos colocar el almacenamiento en nube a largo plazo, que ofrece accesibilidad y flexibilidad, pero es posible que no sea tan duradero y normalmente está asociado a

un costo elevado para su manutención a largo plazo.

Algunos elementos a disposición para la elección de una jerarquía de almacenamiento son las siguientes:

- **Almacenamiento en archivo muerto:** Se considera comúnmente el nivel más bajo de la jerarquía, consiste en emplear medios físicos de almacenamiento de larga duración. Estos medios se utilizan para almacenar recursos digitales que no se acceden con frecuencia pero que se deban de conservar durante largos períodos de tiempo.
- **Almacenamiento en la nube a largo plazo:** Algunos proveedores de servicio en la nube ofrecen soluciones de almacenamiento diseñadas específicamente para la preservación largo plazo de los recursos digitales, con características como el cifrado de extremo a extremo, detección y corrección de errores y redundancia de datos.
- **Archivos digitales en repositorios específicos:** Este nivel implica la utilización de repositorios digitales especializados en la preservación a largo plazo de los recursos digitales, estos pueden estar diseñados bajo el esquema del modelo OAIS como el conocido software libre *archivematica*³ desarrollado por la empresa artefactual.
- **Almacenamiento en base de datos de preservación digital:** Algunas organizaciones optan por el almacenamiento de los recursos digitales en bases de datos especializadas estas pueden ser desarrolladas de manera local bajo los estándares existentes para la preservación digital a largo plazo.
- **Almacenamiento en repositorios de archivos institucionales:** En algunos casos, las instituciones culturales y estatales pueden establecer repositorios de archivos institucionales para la preservación y acceso a largo plazo de los recursos digitales importantes para la organización.



Ejemplo de principio de jerarquía de almacenamiento. Elaboración propia a partir de PIAF

Es importante considerar que el orden de estos niveles de almacenamiento es una estrategia integral de preservación

digital a largo plazo. Esta elección en particular dependerá de diversos factores como la naturaleza de los recursos digitales, requisitos de preservación, el presupuesto disponible y las distintas regulaciones a las que tenga que sujetarse la organización.

3.4. Características de los servicios de almacenamiento

Las características de almacenamiento de servicios se enfocan en satisfacer las necesidades de los usuarios y organizaciones en los términos de almacenamiento y administración de sus datos, entre los más básicos podemos mencionar:

- Capacidad de almacenamiento escalable
- Acceso remoto
- Sincronización de archivos
- Copias de seguridad automática
- Seguridad y cifrado

Los servicios de almacenamiento suelen ofrecer opciones para establecer permisos de acceso granulares, permitiendo a los propietarios de los archivos gestionar quien puede ver, editar o comentar los archivos compartidos. De esta manera se garantiza que la información sensible esté protegida y que solo los usuarios deseados puedan tener acceso.

La colaboración y compartición de archivos es un elemento fundamental normalmente ofrecido para este servicio, permite a los usuarios trabajar de manera conjunta en documentos y archivos de forma remota y en tiempo real, colocar comentarios y realizar notificaciones a los participantes. Esta característica permite el manejo de versiones y restauración de archivos, herramienta útil para rastrear cambios, o recuperar datos perdidos.

El aspecto multiplataforma es también relevante, siendo primordial en la actualidad el acceso a los datos desde los diferentes sistemas operativos, ya sean estos de escritorio (Windows, macOS, Linux) así como los de dispositivos móviles (Android, iOS) permitiendo la mayor parte de sus funcionalidades de manera transparente.

La integración con aplicaciones de terceros es un complemento de apoyo para la integración de la administración de los recursos de los usuarios con sus tareas diarias, algunas integraciones pueden incorporar:

- Clientes de correo electrónico
- Herramientas de colaboración
- Herramientas de edición y diseño
- Gestores de contraseñas
- Herramientas de análisis y visualización de datos

Estas herramientas permiten a los usuarios dar un uso máximo a su almacenamiento, trabajando de manera eficiente y colaborativa de forma integral.

Capítulo 4. Políticas de almacenamiento

Podemos mencionar tres políticas de almacenamiento:

- La aplicación interna
- La mutualización
- La externalización

4.1. Aplicación interna

La aplicación interna ofrece a la organización (una empresa, una institución...) el control completo de la plataforma de almacenamiento y su evolución. Estas libertades de decisión permiten adaptar el conjunto de las disposiciones y de las elecciones a las estrictas necesidades de esta organización.

En contrapartida, esta aplicación requiere importantes recursos humanos y financieros y desarrollo de competencias especializadas.

Esta aplicación interna solo es deseable para grandes empresas o instituciones con suficiente recurso para mantener y actualizar la plataforma a través de los años.

4.2. Mutualización

La mutualización de un servicio de almacenamiento duradero entre varios socios ofrece ventajas importantes:

- Reparto de costos.
- Importantes economías de escala: el costo medio de almacenamiento anual de un GB disminuye significativamente a la función del volumen almacenado.
- Una puesta en común de las competencias que permitan aumentar la seguridad y la fiabilidad del almacenamiento.

Es posible establecer un almacenamiento mutualizado según criterios geográficos o institucionales, independientemente de la lógica del negocio y de los contenidos.

La virtualización brinda toda una gama de herramientas para segmentar la plataforma y obtener el máximo provecho de los recursos mutualizados.

Ejemplos:

- Posibilidad de considerar, a nivel de un departamento francés, la gestión del conjunto de los datos digitales por una sola entidad (expedientes vivos y archivos corrientes tratados por los servicios administrativos y expedientes archivados gestionados por los servicios de archivo).
- Otra posibilidad a nivel interno: almacenamiento seguro y centralizado de los datos y documentos de búsqueda y de las universidades.
- Existencia en los Estados Unidos del San Diego Supercomputer Center (SDSC)⁴ que federa el almacenamiento de datos de 40 universidades, incluida la Universidad de California.
- El estándar «Digital Library Control Systems el proyecto CLARIN (Common Language Resources and Technology Infrastructure)⁵ para los recursos del lenguaje van en esta dirección, pero con un espectro que va mucho más

4 <https://www.sdsc.edu/>

5 <https://www.clarin.eu/>

allá de la cuestión del almacenamiento.

El principal riesgo de la mutualización es la dificultad para encontrar socios comprometidos con la propuesta.

4.3. Externalización (terceros archivadores)

Esta vía permite el recurso a agentes con competencias y medios. Facilita el ahorro de costes de escala, habida cuenta de la importancia de los costes fijos (personal, material). Sin embargo, hay que plantearse la cuestión de la continuidad del proveedor del servicio y de la seguridad de la red que permite transmitir o recuperar datos. Este puede incluir desde el almacenamiento de datos en la nube, utilización de servicios de alojamiento gestionado, contratación de servicios de gestión de datos externo, entre otros.

La norma ISO 14641 titulada “Sistemas de gestión de documentos – Aplicación de la gestión de documentos electrónicos – Principios y requisitos” aborda dentro su contenido las directrices de almacenamiento y preservación, incluyendo la gestión del almacenamiento, copias de seguridad, migración de formatos y medidas de seguridad como la protección de datos.

Capítulo 5. Casos de estudio

Estos casos tienen por objeto resaltar algunos archivos digitales patrimoniales que han logrado con éxito establecer repositorios de preservación a largo plazo, siendo referentes para otras instituciones y organizaciones que han incursionado en proyectos de esta índole:

5.1. Proyecto de archivos de la UNESCO⁶

Establecido en 1992 por la UNESCO (United Nations Educational, Scientific and Cultural Organization) como una iniciativa para la preservación y difusión del patrimonio documental del mundo. Este proyecto se enfoca en la importancia de preservar la diversidad cultural y el patrimonio documental de la humanidad brindando el acceso a beneficio de las generaciones presentes y futuras. Entre estos recursos se encuentran bibliotecas digitales, archivos audiovisuales, documentos digitalizados entre otros materiales de valor cultural e histórico.

5.2. Proyecto de digitalización del patrimonio cultural de Google

La empresa Google ha colaborado con organizaciones como museos, bibliotecas y archivos a través del mundo. Tiene como objetivo difundir el patrimonio cultural a través de la audiencia global de internet. Este proyecto incluye varias iniciativas y programas que se han desarrollado a través de los años, entre estas podemos encontrar:

- **Google Arts & Culture⁷**: Plataforma en línea que brinda acceso a un amplio catálogo de contenido cultural y patrimonial.
- **Google Street View⁸**: Función de Google Maps que permite a los usuarios explorar lugares de interés cultural y patrimonial por medio de imágenes panorámicas en 360 grados, se incluyen museos, bibliotecas, documentación

6 <https://unesdoc.unesco.org/archives>

7 <https://artsandculture.google.com/>

8 <https://www.google.com/intl/es/streetview/>

histórica, entre otros.

5.3. Europeana⁹

Es una plataforma en línea que ofrece acceso a millones de recursos digitales de museos, biblioteca, archivos y galerías de Europa. Fundada en 2008 por la Comisión Europea, alberga de forma gratuita una amplia variedad de recursos digitales, incluyendo audiovisuales y documentos de patrimonio artístico y cultural de Europa.

Bibliografía

Asociación Española de Normalización y Certificación UNE-ISO 14721. (2015). Sistemas de transferencia de datos e información espaciales Sistema abierto de información de archivo (OAIS) Modelo de referencia. Recuperado de Bases de Datos del SIBDI: AENORMÁS.

BANAT-BERGER F., HUC C., DUPLOUY L., *L'Archivage numérique à long terme, les débuts de la maturité?* (Primera obra de síntesis sobre el archivo digital en lengua francesa) Paris, La Documentation française, 2009

BANAT-BERGER F., HUC C., Module 7 - Gestion et archivage des documents numériques. Portail International Archivistique Francophone. 2011. <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques> (Se identifica en el texto como PIAF)

Blumenfeld, J. (18 de Julio de 2017). earthdata.nasa.gov. Obtenido de <https://www.earthdata.nasa.gov/learn/articles/earth-science-data-cloud-eosdis-cumulus-project>

CyberSecurity-Ventures. (2020). cybersecurityventures.com. Obtenido de The 2020 Data Attack Surface Report: <https://cybersecurityventures.com/wp-content/uploads/2020/12/ArcserveDataReport2020.pdf>

IASA. (s.f.). Sistema abierto de archivo de información (OAIS). Obtenido de <https://www.iasa-web.org/tc04-es/618-sistema-abierto-de-archivo-de-informaci%C3%B3n-oais>

9 <https://www.europeana.eu/es>



ARCHIVO NACIONAL
COSTA RICA



UNIVERSIDAD DE
COSTA RICA