# Preservación Digital: problemáticas, estrategias, metadatos, infraestructura y políticas.



## 1. Introducción.

Las colecciones digitales crecen a un ritmo acelerado, como ha sucedido durante los últimos veinte años. Este crecimiento sostenido y, hasta cierto punto, incontrolado, plantea la necesidad de procedimientos que garanticen no sólo la permanencia de las colecciones, sino también que sean consultables y recuperables, independientemente de los cambios tecnológicos.

En este curso se presentan los fundamentos teóricos de la preservación digital y marca la diferencia que existe entre respaldar y preservar, conceptos que suelen emplearse simultáneamente porque tienen aspectos en común, como garantizar a futuro el acceso a los documentos, aunque también marcados contrastes.

### ¿La preservación digital es la preservación de los originales por métodos digitales, o la preservación de los propios materiales digitales?

Ambos enfoques son válidos, existen muchos proyectos de preservación de originales por métodos digitales. Muchas bibliotecas están involucradas en proyectos de digitalización de fondos históricos para mejorar el acceso y, además, contribuir a la preservación del original, ya que el uso de su copia virtual le protege de los efectos nocivos de la manipulación física (Keefer,A. 2003) (Bia, A. 2002). En algunos medios analógicos, tal como las cintas magnéticas, la digitalización ayuda a proteger la calidad de la información (videos por ejemplo) de la degradación natural que sufre el medio en el transcurso del tiempo. El enfoque de preservación de los propios materiales digitales se da debido a la gran fragilidad de los medios de almacenamiento de información digital aunado a los avances rápidos de la tecnología y a la rápida obsolescencia de medios de almacenamiento, hardware y software.

## Definición de respaldo.

El respaldo, también conocido como copia de seguridad, se refiere a la existencia de una réplica de los datos o la información de un sistema, para que éste pueda ser restaurado en caso de fallas o desastres. En este sentido un respaldo es utilizado como un plan de contingencia, para restaurar un equipo de cómputo a un estado operacional luego de un desastre, o bien, para recuperar datos o información que se hayan borrado o corrompido por cualquier causa.

## Definición de preservación digital.

Según (Jones, M. 2001), la preservación digital se refiere a una serie de actividades necesarias y muy bien administradas para asegurar el acceso continuo a los 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. La preservación se clasifica en tres grupos de acuerdo al tiempo:

- **Preservación de duración larga:** Acceso continuo a los materiales digitales o por lo menos a la información contenida en éstos indefinidamente.
- **Preservación de duración media:** Acceso continuo a los materiales digitales aún después de los cambios tecnológicos realizados en un periodo definido de tiempo pero no indefinidamente.
- **Preservación de duración corta:** Acceso a los materiales digitales ya sea por un periodo de tiempo definido o que su uso sea calculado en un periodo de tiempo menor a los cambios tecnológicos.

## Diferencias entre preservación y copias de seguridad

La preservación digital es diferente de las copias seguridad. Lo que se guarda como copia de seguridad en una biblioteca digital son, básicamente, dos cosas: por un lado la información publicada en el servidor (recursos digitales más información de catálogo) y, por otro lado, los recursos digitales en proceso de edición. La preservación digital sin embargo, no se ocupa de respaldar ni los datos del servidor ni el material de trabajo diario, sino de salvaguardar los recursos digitales que necesitaremos en el futuro (Bia, A. 2002). Debido a la limitante en el ancho de banda de red, de muchos usuarios de bibliotecas digitales, comúnmente la información publicada en el servidor está comprimida o sacrifica su calidad para reducir su tamaño y pueda descargarse fácilmente. La información digital, que se desea preservar, debe de ser de la máxima calidad posible para usos futuros. Tal como se indica en (McGray, A.T. 2001)., debe realizarse una separación entre el material para archivo y los derivados para acceso público. Este modelo de biblioteca digital incluye una versión maestra de la biblioteca digital con los recursos de alta calidad (los que se

preservan) y una biblioteca de acceso público con formatos generados automáticamente a partir de la primera.

Si bien las copias de seguridad, al igual que las de preservación, se basan en la redundancia de la información mediante grabaciones periódicas, ni la forma de organizar estas grabaciones ni los tiempos son los mismos. Las copias de seguridad pueden seguir diversos métodos conocidos: copia integral, copia incremental o copias rotativas, por ejemplo, y la periodicidad generalmente es alta (diaria o semanal). En el caso de la copias de preservación, por el contrario, el método suele ser la grabación integral del material una vez y el copiado del mismo una vez al año o cada año y medio en otro soporte nuevo (rejuvenecimiento) (Bia, A. 2002).

## 2. Problemáticas en la preservación digital.

En un estudio realizado en (Preserving our digital heritage, October 2002) se detectaron un conjunto de problemáticas de preservación digital describiendo un grupo general e ilustrando algunas particulares de acuerdo a los siguientes recursos electrónicos: libros, revistas, grabaciones de sonidos, televisión digital, video y páginas WEB. Estas problemáticas junto para la preservación digital, junto con algunas adiciones, se resumen a continuación:

- Nuevos enfoques de seleccionar y catalogar.
- Multiplicidad de formatos.
- Cambios rápidos en la tecnología.
- Obsolescencia de hardware y software
- Problemas legales, sociales y económicos.

**Libros electrónicos**

- Diversidad de iniciativas de estándares.
- Bajo desarrollo en precauciones de seguridad en el mercado.
- Dispositivos de hardware y software propietarios.

**Revistas electrónicas**

- Provistas por ligas a proveedores.
- Contienen artículos repletos con citas a otros recursos secundarios en línea o ligas que probablemente no se preservan.
- ¿Para preservar un artículo hay que preservar todos sus enlaces? ¿tenemos derecho de hacerlo?

**Grabaciones de sonidos**

- Migración de sistemas analógicos a digitales.
- Dependencias de máquinas y medios.
- Obsolescencia de medios.
- Sistemas de almacenamiento masivo.

**Televisión digital y video**

- Migración de sistemas analógicos a digitales.
- Dependencias de máquinas y medios.
- Obsolescencia de medios.
- Demandan sistemas de gran escala de almacenamiento.

**WEB**

- Mortalidad de enlaces demasiado alta.
- Contienen enlaces a otros recursos en línea, de los cuales, algunos probablemente no se preservan.
- ¿Para preservar un documento WEB hay que preservar todos sus enlaces? ¿tenemos derecho de hacerlo?
- ¿Cómo definir los límites en los enlaces de un servidor web?

# 3. Respaldo y recuperación

Los sistemas de cómputo que contienen las colecciones digitales, están expuestos a riesgos latentes. Pueden verse interrumpidos en su servicio, debido a alteraciones en la electricidad, el hardware, el software y la red, así como fallas humanas, desastres naturales y ataques informáticos como virus y sabotaje, entre otras.

El riesgo a que están expuestos nuestros sistemas informáticos es inminente, por tal motivo es necesario contar con un plan de contingencia adecuado que garantice la recuperación de la información así como la disponibilidad del sistema informático que la gestiona. De acuerdo con (Hernández, I. 2005) para la elaboración del plan de contingencia es necesario lo siguiente:

- Identificar y priorizar los procesos y los recursos indispensables;
- Analizar el riesgo y el impacto por la pérdida de la información;
- Evaluar recomendaciones de protección;
- Contar con estrategias y alternativas de recuperación;
- Establecer los equipos de trabajo y las funciones de cada persona;
- Ejecutar simulacros del plan de contingencia;
- Elaborar un manual de contingencia, y
- Retroalimentar el plan.

Para llevar a cabo un plan de contingencia, es recomendable realizar algunas de las siguientes actividades:

- Seleccionar el medio de almacenamiento secundario;
- Determinar la frecuencia de realización de copias de seguridad;
- Determinar el volumen de la información a respaldar, y
- Determinar días y horario en que deben realizarse los respaldos.

Además de permitir la identificación de la mejor manera de recuperar la información en caso de desastre, una *estrategia de recuperación* es una guía para el desarrollo de los procedimientos mismos de recuperación.

## 3.1 Respaldo tradicional

El respaldo tradicional consiste en copiar los datos o la información de un sistema, a un medio de almacenamiento secundario, como cinta, CD y DVD, entre otros, con el fin de que pueda ser restaurado en caso de fallas o desastres. Su periodicidad puede ser diaria, semanal o mensual y difícilmente menor a un día. Para realizar las copias, los métodos a seguir pueden ser los siguientes:

- *Copiar sólo los datos*. No provee las facilidades para recuperar el entorno operacional, que proporcionan los programas de aplicación para acceder a los mismos;
- *Copia completa*. Incluye una copia de datos y programas, que permite restaurar el sistema hasta el momento anterior a la copia;
- *Copia incremental*. Solamente se almacenan las modificaciones realizadas después de la última copia de seguridad. Debe mantenerse la copia original para restaurar posteriormente el resto de las copias, y
- *Copia diferencial*. Es similar a la incremental, pero en lugar de copiar las modificaciones, son almacenados los archivos completos que han sido modificados. También se necesita la copia original.

## 3.2 Respaldo con tecnología RAID

En el mejor de los casos, el sistema de respaldo tradicional se aplica todos los días, comúnmente por la noche, cuando disminuye la carga de trabajo del servidor. Esto significa que si se presenta un incidente en el transcurso del día o, en las circunstancias más adversas, por la tarde, no sería posible recuperar el trabajo realizado. Para muchas empresas esto puede representar grandes pérdidas financieras. En el caso de los bancos, éstos simplemente no pueden perder las transacciones realizadas a lo largo del día. Para este tipo de contingencias la solución tecnológica es el uso del RAID (*Redundant Array of Inexpensive Disks* o Conjunto redundante de discos baratos y, actualmente, *Redundant Array of Independent Disks* o Conjunto redundante de discos independientes).

En informática, el acrónimo RAID se refiere a un sistema de almacenamiento en el que se usan múltiples discos duros, entre los que son distribuidos o replicados los datos. Dependiendo de su configuración, a la que suele denominarse “nivel”, los beneficios de un RAID con respecto a un único disco, son:

- Mayor integridad.
- Tolerancia a fallos.
- Rendimiento y capacidad.

En sus orígenes, la principal ventaja de RAID radicaba en su capacidad de combinar varios dispositivos de bajo costo con una tecnología más antigua, para dar como resultado un conjunto que ofrecía mayor capacidad, fiabilidad,

velocidad, o una combinación de éstas, que un solo dispositivo de última generación y costo mayor.[1]

En el nivel más simple, RAID combina múltiples discos en una sola unidad lógica. Entonces en lugar de ver diferentes discos, el sistema operativo sólo ve uno. El RAID agrupa dos o más discos duros ofreciendo una forma más avanzada de respaldo ya que[2]:

- Es posible mantener copias en línea (redundancia).
- Agiliza las operaciones del sistema, sobre todo en bases de datos.
- El sistema es capaz de recuperar información, sin la intervención de un administrador.

Hablar del nivel o la configuración del RAID, es referirse a la arquitectura que determina la redundancia y cómo están distribuidos los datos a través de los discos duros del arreglo. Existen varias configuraciones del RAID, sin embargo, los cuatro tipos que prevalecen en muchas arquitecturas, son: RAID-0, RAID-1, RAID-3 y RAID-5.

Cabe aclarar que para la implantación de la tecnología del RAID, se requiere un presupuesto mayor. Por este motivo es una tarea importante de las instituciones analizar y evaluar, en función de sus recursos financieros y necesidades, la tecnología a utilizar.

# 4. Estrategias para la preservación digital

Existe un conjunto de estrategias de preservación digital que deben aplicarse durante el tiempo de vida de un recurso digital para garantizar su preservación en el corto, mediano y largo plazo. A continuación se describen cada una de ellas:

**Preservación de la tecnología**: consiste en preservar el ambiente tecnológico para visualizar y editar el contenido digital, incluyendo software y hardware, como por ejemplo: sistemas operativos, programas de visualización, periféricos de lectura y escritura de medios de almacenamiento secundario.

Tomada de: http://www.iimas.unam.mx/iimas/pagina/es/19/quienes-somoses

[1] http://es.wikipedia.org/wiki/RAID consultado en octubre de 2007

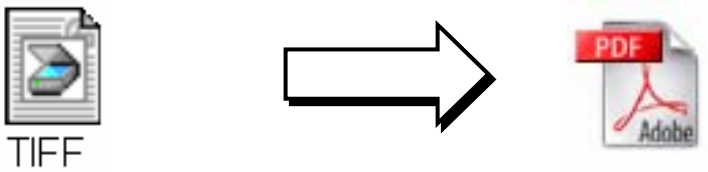
[2] http://www.monografias.com/trabajos14/respaldoinfo/respaldoinfo.shtml consultado en octubre de 2007

**Migración:** Se refiere a superar la obsolescencia tecnológica al transferir o adaptar el contenido digital de una generación de hardware y software hacia otra generación. Tiene la desventaja de ocasionar pérdidas en la información tras migraciones sucesivas.

Ejemplo:

**Reformateo:** Se refiere a cambiar el contenido digital de un formato a otro.

Ejemplo:

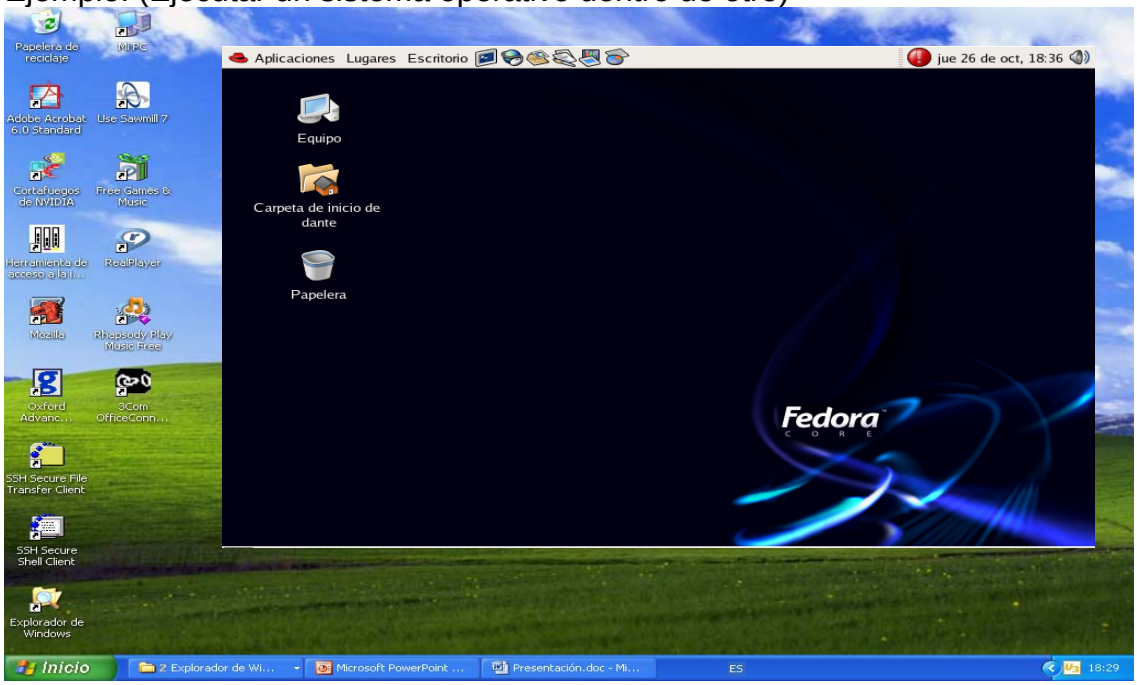



**Rejuvenecimiento:** Se refiere a copiar el contenido digital de medio de almacenamiento a otro nuevo del mismo tipo. O bien, escribir cada determinado tiempo el contenido digital a un medio nuevo para evitar que el contenido se pierda a causa de la degradación natural del medio por el transcurso del tiempo.

Ejemplo:

**Emulación:** Según (Waugh, A. 2000), la emulación permite que el software original sea usado sin necesidad de que el sistema original que lo ejecutaba siga existiendo. La emulación obliga a preservar una cantidad importante de información. Una solución de emulación por hardware, por ejemplo, implica la preservación del emulador, el sistema operativo, la aplicación y los datos.

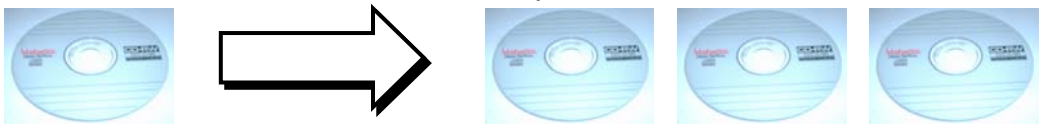
Ejemplo: (Ejecutar un sistema operativo dentro de otro)



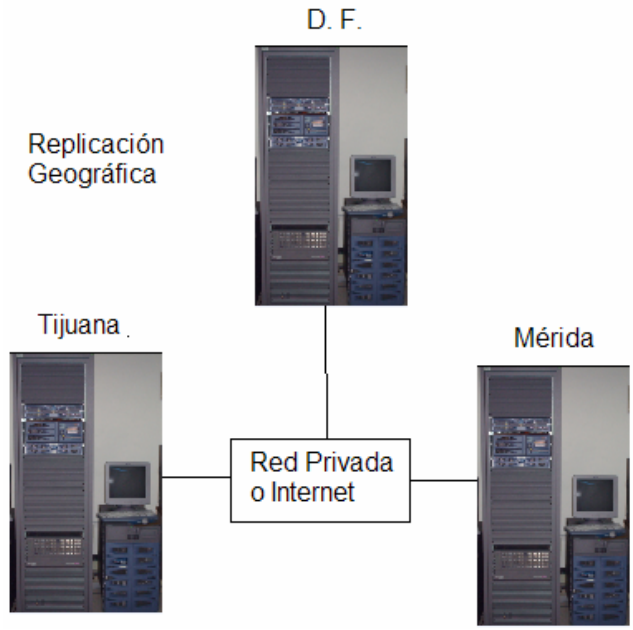

**Replicación:** Se refiere al hecho de mantener una o mas copias de un mismo contenido digital.

Ejemplos:

Maestro



Copias

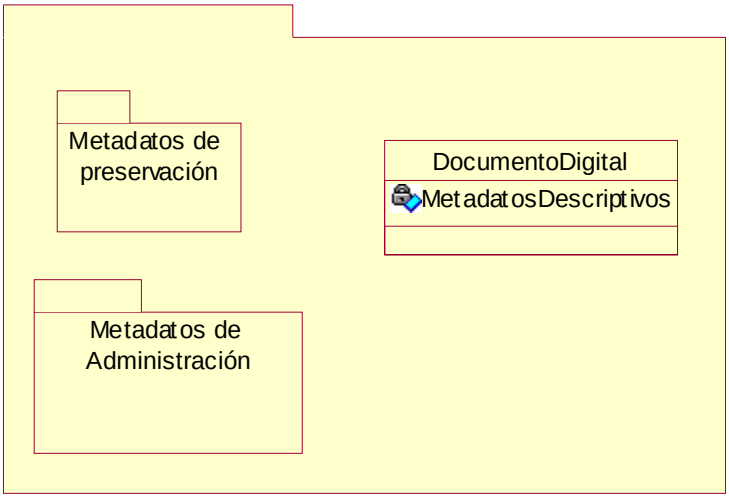


**Estandarización:** Se refiere al hecho de utilizar algún formato estándar para la representación del material digital. Esto garantiza un mejor soporte de herramientas para administrar el material digital, una mayor duración del formato y una mejor migración ante los cambios tecnológicos.

Ejemplo:

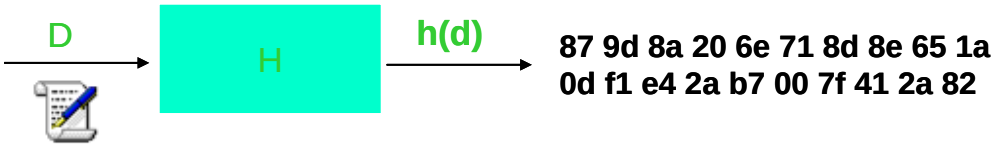

**Encapsulado:** Según (Waugh, A. 2000), se refiere a empaquetar la información que se desea preservar junto con un diccionario de datos (metadatos descriptivos) y mantenerlos en una única localización. Además de la emulación existen los otros factores clave para la preservación de duración larga: Auto documentación (es decir, la capacidad de entender y decodificar la información preservada sin hacer referencia a documentación externa), Auto suficiencia (minimización de dependencias con respeto a sistemas, datos o documentación), documentación de contenido (habilidad para que un usuario futuro encuentre o implante el software para visualizar la información preservada), preservación de organización (habilidad para almacenar la información que le permita a la organización el uso eficiente de la información preservada.

Ejemplo: Paquete de Información



**Autenticidad:** Se refiere al hecho de asegurar la integridad de la información digital. Existen muchas causas por las cuales se puede corromper la información digital: virus, negligencias, fallas de los medios de almacenamiento, ataques informáticos maliciosos, etc. Para asegurar la autenticidad se propone utilizar firmas digitales sobre la información digital.

Ejemplo:




D - Documento Digital

H - Función Hash

h(d) - Hash del Documento Digital

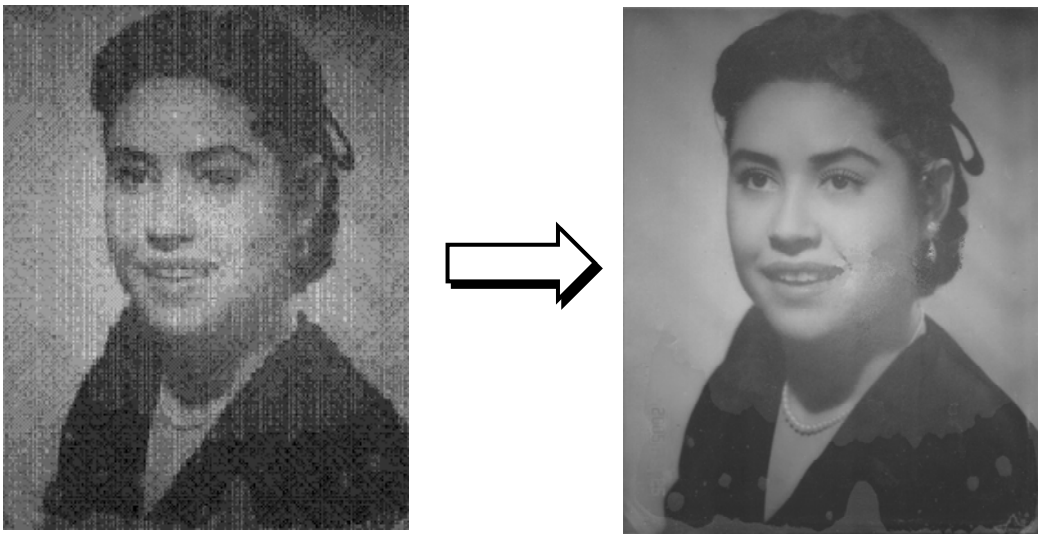

Firma Digital

Certificado Digital

Autoridad Certificadora

**Arqueología digital:** Es el proceso de recuperar la información a partir de medios de almacenamiento digital dañados o antiguos.

Ejemplo: restauración de imágenes



**Cuidado duradero**. El cuidado duradero debe ser visto como una estrategia continua para asegurar que los documentos digitales se encuentren en óptimas condiciones. En el cuidado de una colección, los archivos deben almacenarse en medios y ubicaciones no sólo seguros, sino también confiables. Además, deben manipularse con base en las pautas de aceptación internacional, orientadas a optimizar su expectativa y la calidad de duración.

Ejemplo: Modo de colocación y almacenamiento


## 5. Modelo de referencia OAIS

El modelo de referencia OAIS (Open Archival Information System) ha sido publicado como una recomendación del *CCSDS* (*Consultative Committee for Space Data Systems*) y como norma *ISO14721:2003*. *OAIS* enfoca su actividad en la preservación a largo plazo de la información en formato digital, como garantía de que será accesible en el futuro.

### Introducción

Un sistema de información debe considerar el hardware, el software y los recursos humanos necesarios para la adquisición, preservación y difusión de la información. *OAIS* consiste en un modelo lógico que abarca todas las funciones de un repositorio digital, señalando la forma en que los objetos digitales deben ser preparados, enviados a un archivo, almacenados durante largos períodos, conservados y recuperados (Silió, T. 2005).

El modelo de referencia OAIS se ha convertido en el concepto más reconocido de un sistema que involucre preservación digital. El documento completo contiene 148 páginas y esta estructurado en seis secciones (Introducción, Conceptos de OAIS, Responsabilidades en OAIS, Modelo detallado, Estrategias de preservación, Interoperabilidad de archivos) y seis anexos (Ejemplos de archivos existentes, Relaciones con otros estándares y esfuerzos, Guía breve del Lenguaje de Modelado Unificado, Referencias Informativas, Un modelo para uso de software en representación de información, Vista funcional compuesta) que proporcionan información vital a las organizaciones que tratan de implantar un sistema de archivado digital acorde con OAIS.

OAIS trata de identificar las responsabilidades y componentes de un sistema de archivado incluyendo:

- Las funciones de las personas e instituciones que interactúan con un archivo. Estos se ven en el modelo como **Productor**, **Administrador** y **Consumidor**.

- Objetos digitales que son manejados por OAIS y que se denominan **paquetes de información.**

- Seis funciones de nivel alto que aparecen en el modelo como Ingesta, Administración de Datos, Almacén de Archivos, Acceso, Planeación de Preservación y Administración que representan treinta y tres funciones de nivel bajo.

## Entidades de OAIS

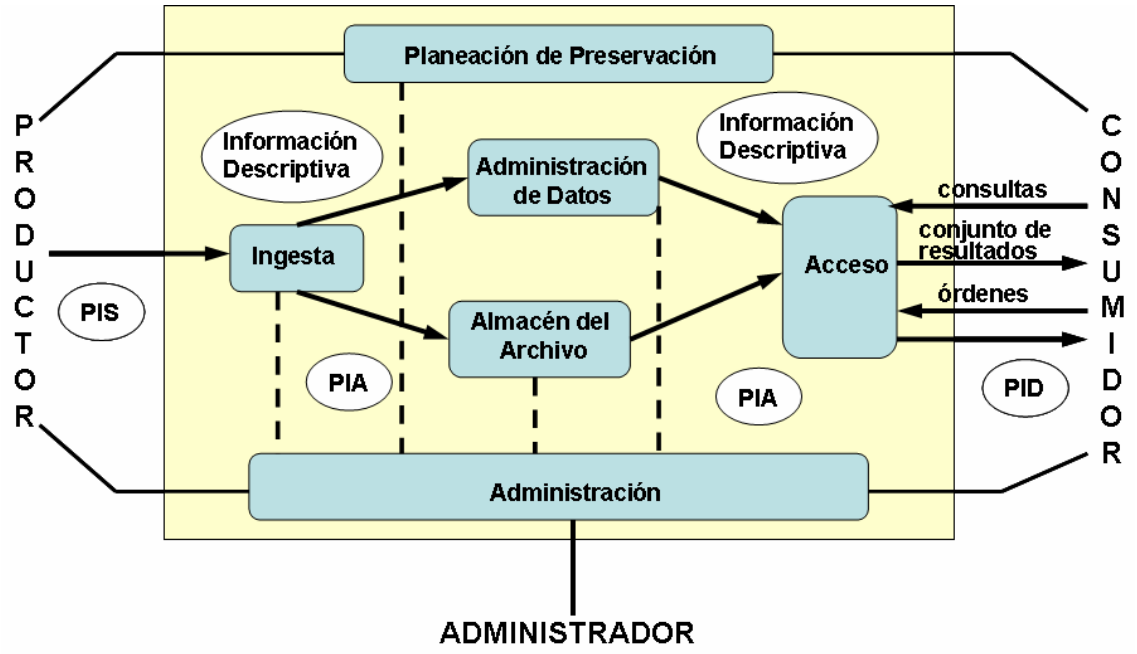
El diagrama OAIS ilustra las relaciones entre las funciones. Aunque se ve como un diagrama de arquitectura no lo es. Los rectángulos identifican grupos de funciones relacionadas y no componentes en una implantación. En el mundo real las funciones no necesitan estar en el mismo servidor o en la misma organización. Los grupos pueden estar separados y su funciones distribuidas en muchas configuraciones.

En el exterior de OAIS se encuentran los Productores, Consumidores y Administradores:

- **Productor** es la función desarrollada por aquellas personas que proveen, al sistema, la información que debe ser preservada.

- **Administrador** es función desarrollada por aquellas personas que definen las políticas globales como un componente del amplio dominio de políticas. El control de la administración de OAIS es solo una de las responsabilidades del Administrador. El Administrador no esta involucrado en las operaciones diarias del archivo. Las operaciones diarias sobre el archivo son responsabilidad del la entidad funcional Administración.

- **Consumidor** es la función desarrollada por aquellas personas que interactúan con los servicios de OAIS para encontrar y obtener la información preservada que sea de su interés.




**Entidades Funcionales de OAIS**

## Modelo Funcional de OAIS

El modelo de referencia OAIS de compone de seis entidades funcionales y sus interfaces relacionadas. En la figura general que muestra el modelo solo se muestran los flujos de información más importantes. Las líneas que conectan a las entidades identifican rutas de información, sobre las cuales, la información fluye en ambas direcciones. Se utilizan líneas discontinuas para evitar confusión.

**Ingesta:** Esta entidad proporciona los servicios y funciones para aceptar los Paquetes de Información Sometida (PISs) de los productores (o de los elementos internos bajo el control de la Administración) y prepara el contenido para el manejo y almacenamiento en el archivo. Las funciones de Ingesta incluyen la recepción de PISs asegurando su calidad y generando el Paquete de Información de Archivado (PIA) el cual cumple con los estándares de documentación y formateo de datos, extrae información descriptiva de los PIAs para inclusión en la base de datos del archivo. Coordina actualizaciones en Almacén del Archivo y Administración de Datos.

**Almacén del Archivo:** Esta entidad proporciona los servicios y funciones para el almacenamiento, mantenimiento y recuperación de PIAs. Sus funciones incluyen la recepción de PIAs de Ingesta y agrega estos para almacenamiento permanente administrando una jerarquía de almacenamiento, refrescando los medios, sobre los cuales, los contenedores de los archivos son almacenados ejecutando rutinas y verificación de errores brindando capacidades para recuperación de desastres y AIPs para satisfacer las órdenes.

**Administración de Datos:** Esta entidad proporciona los servicios y funciones para poblar, mantener y acceder tanto información descriptiva que identifica y documenta contenedores de archivos como datos administrativos para manejo del archivo. Sus funciones incluyen la administración de la base de datos del archivo (manteniendo las definiciones del esquema y vistas e integridad referencial), ejecutando actualizaciones de la base de datos (cargando información descriptiva nueva o datos administrativos del archivo), ejecutando consultas sobre datos para administración de datos y generar conjuntos de resultados y producir los reportes respectivos.

**Administración:** Esta entidad proporciona los servicios y funciones para la operación global del sistema de archivo. Las funciones de administración incluyen la solicitud y negociación de los acuerdos de sometimiento con los productores, auditan los sometimientos para asegurar que cumplen los estándares de archivo. Mantiene la administración de la configuración del software y hardware del sistema. Proporciona también funciones de ingeniería del sistema para monitoreo y mejoramiento de las operaciones del archivo, el inventario, reportes y migración/actualización del contenido del archivo. Es responsable de establecer y mantener políticas y estándares del archivo brindando soporte a los usuarios y habilitando las solicitudes almacenadas.

**Planeación de Preservación:** Esta entidad proporciona los servicios y funciones para monitoreo del ambiente de OAIS, brindando recomendaciones que aseguren que la información almacenada en el sistema de archivado permanezca accesible a la comunidad de usuarios diseñada en un tiempo bastante prolongado, aún si el ambiente de computación original se vuelve obsoleto. Las funciones incluyen la evaluación del contenido del archivo y recomendaciones periódicas de actualización de la información del archivo para migar los contenedores actuales de los archivos, desarrollando recomendaciones sobre políticas y estándares de archivo y monitoreando cambios en el ambiente tecnológico y en los requerimientos de servicios y base de conocimiento de la comunidad de usuarios diseñada. Planeación de Preservación también diseña modelos de paquetes de información brindando asistencia y revisión del diseño para especializar estos modelos en PISs y PIAs para sometimientos específicos. Planeación de preservación también desarrolla planes de migración detallada, prototipos de software y planes de pruebas para liberar implantaciones de los objetivos de migración de Administración.

**Acceso:** Esta entidad proporciona los servicios y funciones para soporte a los consumidores en la obtención de la existencia, descripción, localización y disponibilidad de información almacenada, en el sistema de archivado, permitiendo a los consumidores solicitar y recibir productos de información. Las funciones de acceso incluyen la comunicación con los consumidores para recibir solicitudes aplicando controles que limitan el acceso a la información protegida, coordinando la ejecución de solicitudes para que se completen exitosamente, generando respuestas (PIDs, conjuntos de resultados, reportes) para entrega a los consumidores.

# 6. Estándares de metadatos en preservación digital

Los metadatos de preservación son un conjunto de datos estructurados que permiten codificar, como parte del mismo documento digital, información relacionada con su preservación, es decir, en qué formato se generó, con qué compresión, calidad, etcétera. Asimismo, a través de ellos es posible identificar, describir, clasificar y localizar los documentos digitales que se preservarán.

A continuación se enlistan algunos proyectos e iniciativas, en los que se han desarrollado estándares de metadatos para la preservación digital:

RGL/ OCL Working Group on Preservation Metadata
http://www.rlg.org./preserv/presmeta.html

NEDLIB (Networked European Deposit Library)
http://nedlib.kb.nl/results/D4.2/D4.2.htm#_Toc494249797

Metadata for Digital Preservation : the Cedars Project Outline Specification
http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html

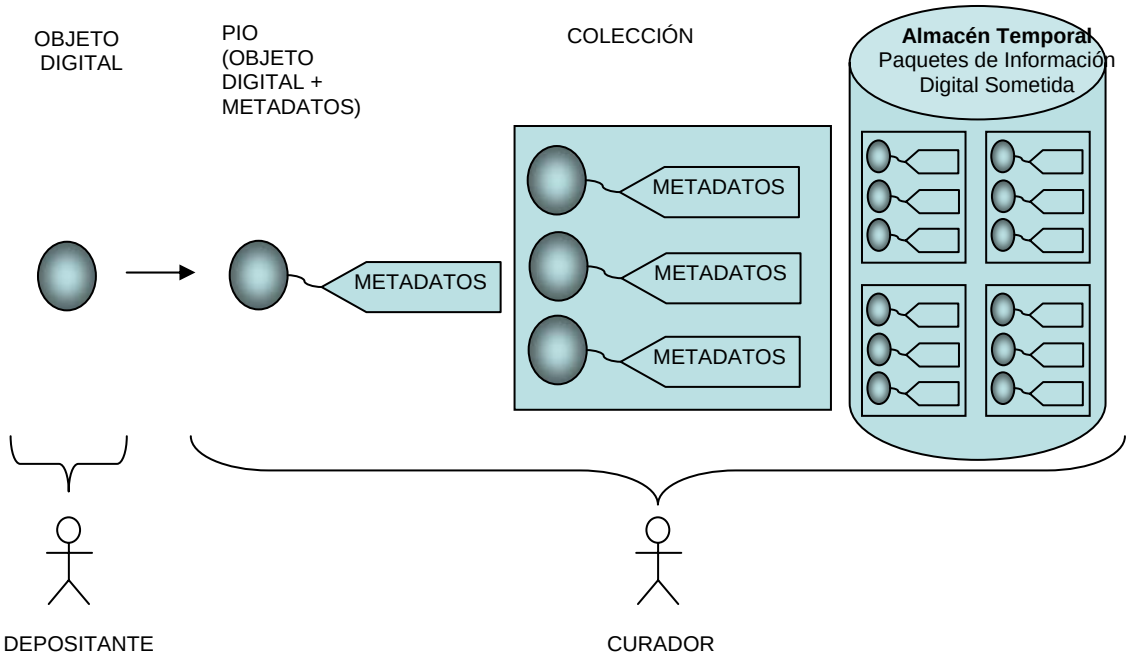
National Library of Australia, Preservation Metadata for Digital Collections
http://www.nla.gov.au/preserve/pmeta.html

# 7. Infraestructura Tecnológica

Todas las figuras siguientes muestran los requerimientos mínimos de la infraestructura tecnológica para el manejo de la preservación digital. El depositante representa la comunidad de usuarios que producen los documentos digitales, que se ofrecen a través del servidor de publicación. El administrador de la colección o curador es quien valida los documentos digitales que se someten a depósito en el servidor de publicación. Otra de sus funciones es realizar el depósito de los documentos digitales de publicación y preservación, en el servidor de preservación.

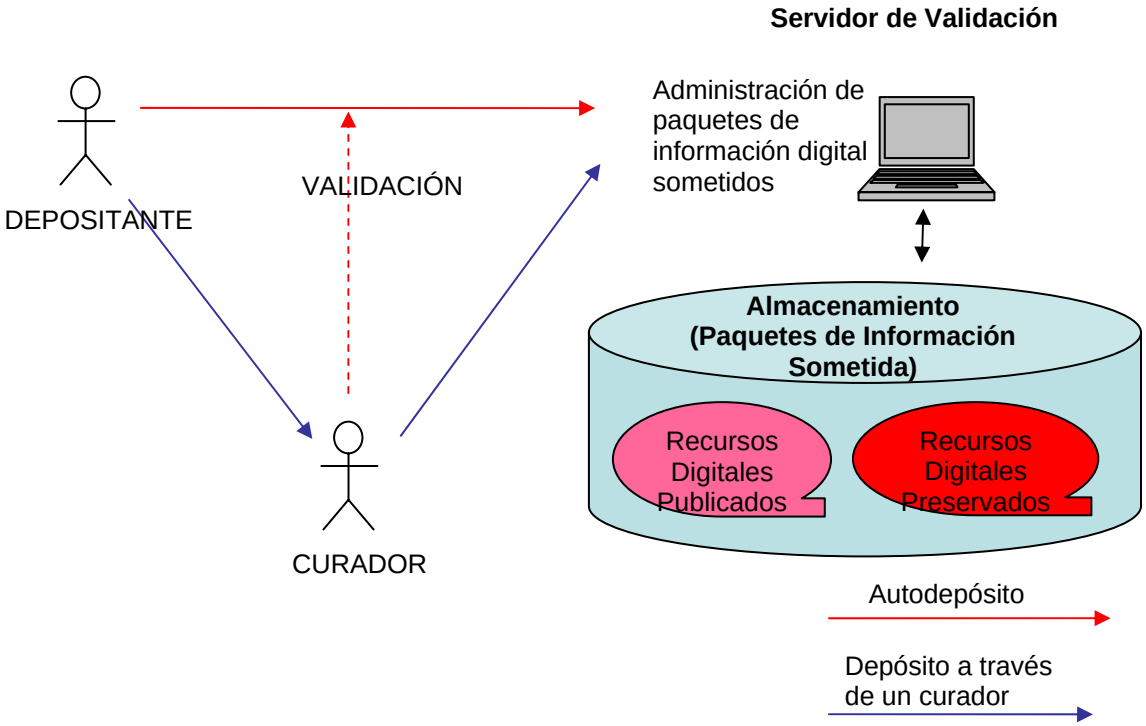
## Construcción de Paquetes de Información Digital

Tal y como se especifica en el modelo OAIS, se debe primeramente construir el paquete de información sometida (PIO) al integrar el objeto digital y sus metadatos descriptivos. Se pueden agrupar los PIOs por colecciones para mantener una buena organización de contenido. Las colecciones pueden ser por tema, tipo de documento, departamento, etc.



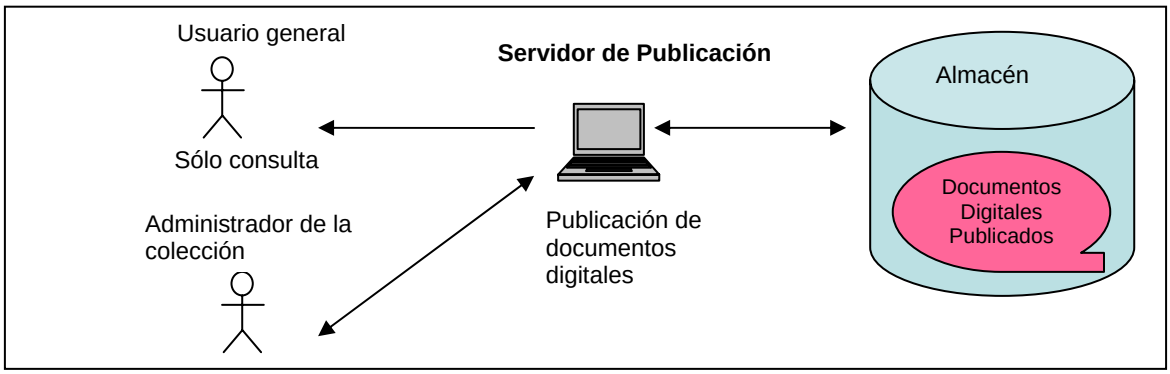

## Validación de Paquetes de Información Digital

A fin de mantener un control sobre el almacenamiento de PIOs, el depositante incorpora el objeto digital en el servidor de validación y registra los metadatos descriptivos o de otro tipo. Este tipo de depósito es conocido como depósito directo. El curador recupera el PIO para validarlo y para que posteriormente puedan depositarse los servidores de publicación y preservación. El depósito indirecto se realiza cuando el depositante entrega el objeto digital junto con sus metadatos al curador para que este proceda a validarlos e introducirlos en los servidores correspondientes.




## Arquitectura del servidor de publicación

La figura siguiente muestra la arquitectura del servidor de publicación. Con el usuario general se representa la comunidad que puede consultar los documentos digitales, que se ofrecen a través del servidor de publicación.

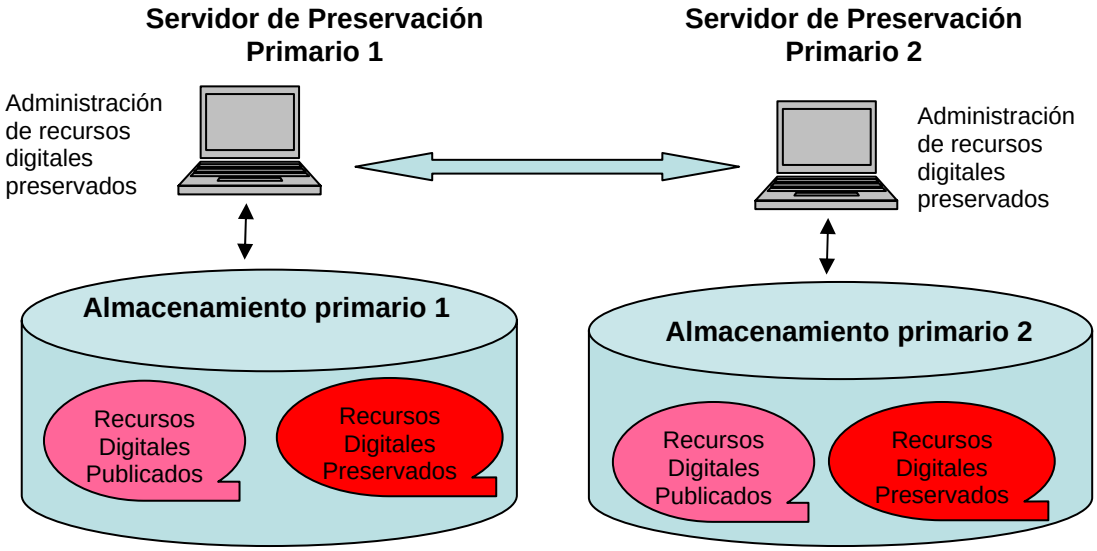



## Replicación de servidores primarios de preservación

Para asegurar el acceso continuo a los recursos digitales preservados es necesario contar como mínimo con un servidor primario de preservación. El acceso a este servidor será posible únicamente a través del servidor de publicación y sólo cuando no se pueda recuperar por medio de su respaldo un documento digital publicado. Para tener una alta disponibilidad y mejores garantías para proveer el acceso continuo se recomienda que este servidor se

encuentre replicado geográficamente. El acceso al servidor de preservación será posible únicamente por:

- El servidor de publicación que desea un documento digital publicado y sólo cuando no se pueda recuperar por medio de su respaldo.
- El servidor de validación cuando requiera incorporar contenido nuevo.



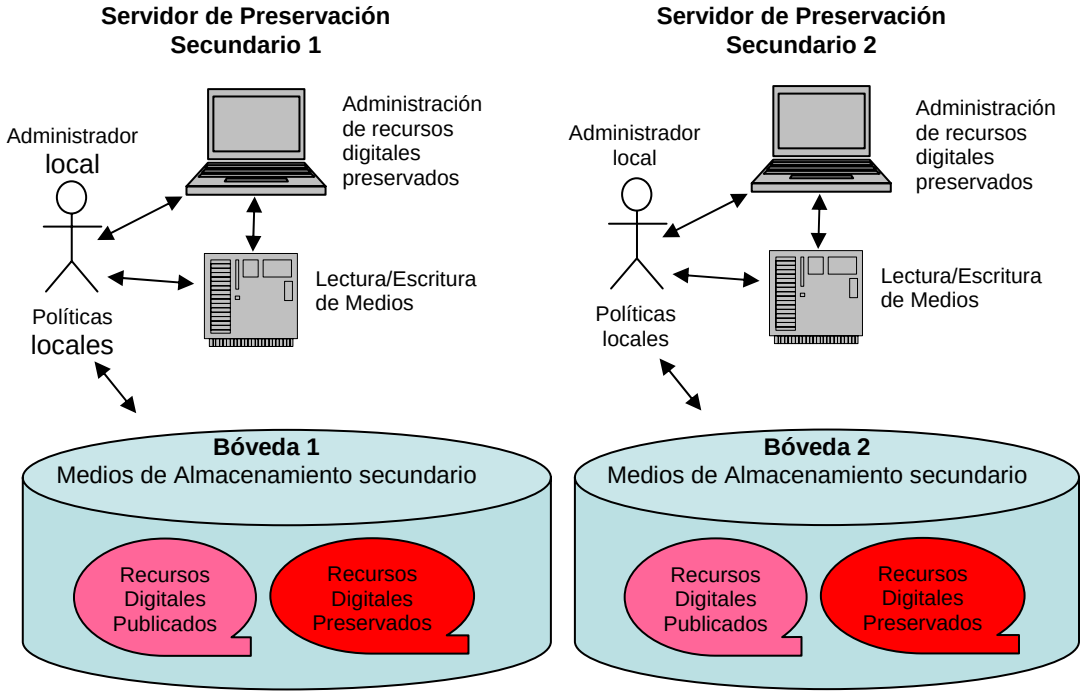

## Replicación de servidores secundarios de preservación

En el resguardo de los medios de almacenamiento secundario, como es el caso de los documentos digitales preservados, es imprescindible contar con una bóveda. Si lo permiten los recursos, para ofrecer una mayor disponibilidad, además de mejores garantías de preservación ante fallas y desastres naturales, es ineludible tener una réplica de la bóveda localizada geográficamente.

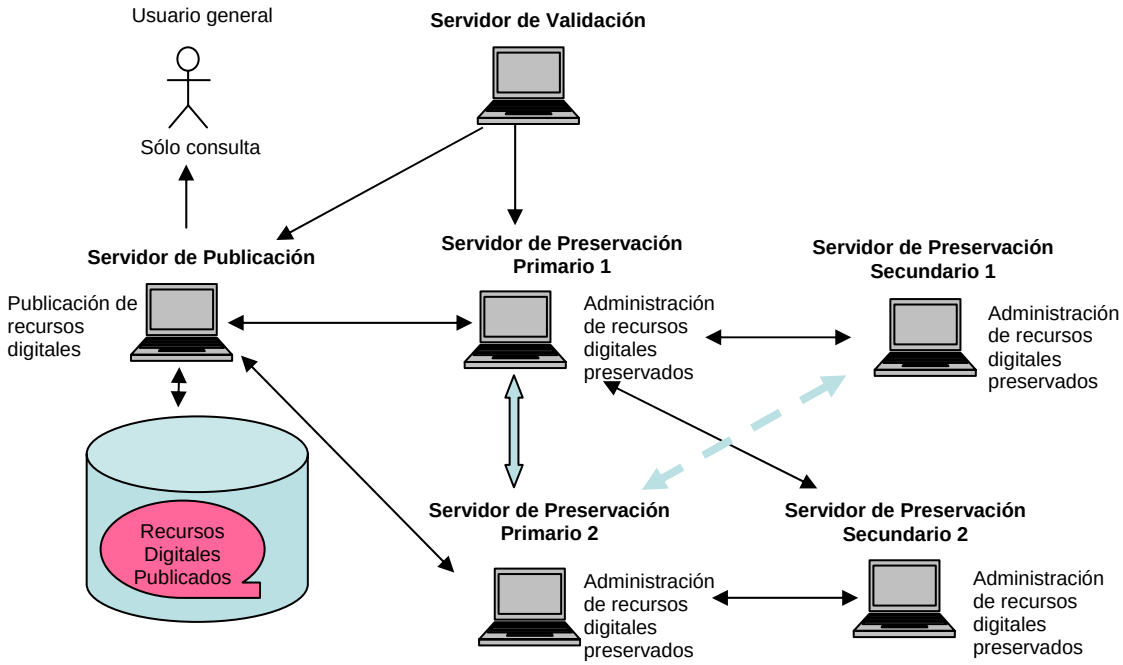
Debe haber por lo menos un administrador local o curador, que mantenga organizados los medios de almacenamiento en la bóveda y tenga la autoridad para incorporarlos o retirarlos. Su obligación será cumplir las políticas y los procedimientos durante la administración de los documentos digitales publicados y preservados. Cuando le sea solicitado, el Administrador local podrá depositar uno o más documentos digitales publicados o preservados en el servidor secundario de preservación. El acceso al servidor secundario de preservación será posible únicamente a través de cualquier servidor primario de preservación y en los siguientes casos:

- No se pueda recuperar un documento digital publicado o preservado por medio del respaldo.
- El servidor primario de preservación requiere incorporar contenido nuevo.

Otra de las actividades del Administrador local, será la generación de los medios de almacenamiento secundario de los documentos digitales publicados y preservados, que hayan sido depositados en el servidor de preservación, por el servidor primario de preservación.




## Arquitectura tecnológica



Para los servidores de publicación y preservación se requieren, en general, equipos de gran capacidad, alta disponibilidad, escalabilidad, adaptabilidad y bajo costo.

Se hace necesaria una gran capacidad para:

- Brindar el servicio;
- Detectar y eliminar cuellos de botella;
- Tener velocidad de respuesta en los equipos, y
- Favorecer la comunicación de la red.

Una alta disponibilidad es deseable para:

- Ofrecer un servicio continuo;
- Tener interrupciones breves;
- Operar con equipos o componentes redundantes (es decir, duplicados, replicados o repetidos);
- Tolerar fallas, y
- Asegurar una recuperación automática frente a un problema.

Las características de los equipos, relacionadas con la escalabilidad y la adaptabilidad, les permitirá crecer conforme a la demanda.

# 8. Políticas y procedimientos

Dentro del contexto relacionado con el respaldo y la preservación, las políticas y los procedimientos son un conjunto de métodos que, aplicados sistemáticamente, sirven de apoyo en la realización del respaldo, el resguardo, la recuperación y la preservación de un contenido digital. Cada institución determina su propio conjunto de políticas y procedimientos, aplicables sólo dentro de ella. Las siguientes son algunas políticas y procedimientos generales para el respaldo y la preservación de documentos digitales, aplicables en cualquier proyecto de digitalización.

## Políticas y procedimientos de respaldo y preservación

1. Manejar con mucho cuidado los medios de almacenamiento.

2. Cumplir con las especificaciones del fabricante para el cuidado de los medios de almacenamiento, como las condiciones climáticas: humedad, calor, polvo, etcétera.

3. Acomodar los medios de almacenamiento en forma vertical.

4. No colocar objetos sobre los medios de almacenamiento.

5. Verificar la integridad del contenido almacenado en el dispositivo de almacenamiento secundario, cada vez que se realice una copia de la información.

6. Verificar periódicamente el funcionamiento correcto del dispositivo periférico, para la generación de copias de los datos.

7. Establecer reglas y procedimientos para la integración de metadatos.

8. Validar que los documentos digitales a ingresar se encuentren en un formato estándar.

## Políticas y procedimientos de respaldo

1. Los respaldos deben hacerse en el horario de menor uso del servidor de publicación.

2. Se recomienda tener una copia del contenido digital cerca del servidor de publicación y otra, lejos.

3. Retirar el medio de almacenamiento secundario de la unidad de lectura y grabación, cuando haya concluido el proceso de respaldo.

4. Cumplir con los periodos de respaldo indicados en el plan de seguridad y contingencia.

## Políticas y procedimientos de preservación

1. Contar con un mínimo de dos bóvedas replicadas geográficamente para garantizar la preservación de los medios de almacenamiento contra desastres naturales.

2. El acceso a las bóvedas debe estar restringido a un número limitado y bien definido de personas.

3. Contar con un mínimo de dos servidores primarios de preservación, replicados geográficamente, para garantizar el acceso continuo de los recursos electrónicos preservados y como un mecanismo adicional preservación contra desastres naturales.

4. La administración de los servidores primarios de preservación debe realizarse preferentemente de forma local, o en su defecto, limitado a un máximo de tres sitios.

5. El acceso a los servidores secundarios de preservación solo puede realizarse por los servidores primarios de preservación ya sea para depósito u obtención de paquetes de información.

6. La consulta de los servidores secundarios de preservación solo puede realizarse por el servidor de publicación de recursos electrónicos.

7. Refrescar los medios una vez al año.

8. Evitar el uso de los maestros de preservación.

9. Por cada maestro de preservación generar un mínimo de dos copias.

10. Por cada maestro de publicación generar un mínimo de dos copias.

11. Asignar un límite de vida a cada recurso electrónico.

12. Verificar semestralmente cambios tecnológicos en los formatos de almacenamiento.

13. Verificar semestralmente cambios tecnológicos en software y hardware que impacten en la obsolescencia del software y hardware en uso.

14. Verificar semestralmente implantación de estándares nuevos en representación e intercambio de información, metadatos descriptivos y de preservación.

15. Verificar ya sea manualmente o con programas la sincronía de los recursos digitales almacenados en los servidores primarios de preservación.

16. Verificar ya sea manualmente o con programas computacionales la integridad de los recursos digitales almacenados en los servidores primarios de preservación.

17. Realizar preferentemente con apoyo de programas computacionales, la migración o reformateo de los recursos electrónicos cada vez que lo sugieran los cambios tecnológicos o los estándares nuevos.

18. Validar que los paquetes de información sometidos se encuentren bien documentados (metadatos descriptivos, de preservación y administración) y sean autosuficientes.

# 9. Proyectos e iniciativas para la preservación digital

Por sí solas las soluciones técnicas no son suficientes para asegurar la duración prolongada de los documentos digitales. Para lograr soluciones plenas y satisfactorias, se requiere la integración de aspectos técnicos y administrativos: recursos humanos, capacitación, requisitos financieros, criterios de selección, metadatos de preservación, etcétera.

Para una administración efectiva de las colecciones digitales, se debe desarrollar y seguir un plan de gestión en los proyectos, que permita evaluar los requisitos de preservación y el acceso a largo plazo. Simultáneamente, deben ser identificados los costos y los beneficios, además de estimarse los

riesgos. A continuación se muestran algunos de los proyectos e iniciativas más sobresalientes en el ámbito de preservación digital:

NEDLIB (Networked European Deposit Library)
http://nedlib.kb.nl/

Cedars (curl exemplars in digital archives)
http://www.leeds.ac.uk/cedars/

CAMILEON (Creative Archiving at Michigan and Leeds Emulation the Old On the New)
http://www.si.umich.edu/CAMILEON/

DPC (Digital Preservation Coalition)
http://www.dpconline.org/

PANDORA (Preserving and Accessing Networked Documentary Resources of Australia)
http://pandora.nla.gov.au/

NDIIPP (The National Digital Information Infrastructure and Preservation Program) Library of Congress
http://www.digitalpreservation.gov/

PADI (Preserving Access to Digital Information) National Library of Australia
http://www.nla.gov.au/padi

## Referencias

Bia, Alejandro., & Sánchez, Manuel. (Septiembre 2002). "Desarrollo de una política de preservación digital: tecnología, planificación y perseverancia". *Jornadas sobre Bibliotecas Digitales*. Retrieved from http://mariachi.dsic.upv.es/jbidi/jbidi2002/Camera-ready/Sesion1/S1-4.pdf

Hernánez Zapardiel, Ignacio José. (Diciembre 2005) Métodos y Políticas de Respaldo (backup) en Planes de Contingencia. Universidad Politécnica de Madrid, España. Retrieved from www.criptored.upm.es/guiateoria/gt_m001l.htm

Jones, M., & Beagrie N. (2001). Preservation Management of Digital Materials. *British Library Cataloging in Publication Data*.

Keefer, Alice., & Gallart, Núria. (2003). La preservación digital y las universidades: el estado de la cuestión. *8as Jornadas Españolas de Documentación*. URI: http://hdl.handle.net/10760/6780

McGray, A.T., & Gallagher M.E. (2001). Principles for Digital Libraries Development. *Communications of the ACM,* 44, 49-54.

Preserving our digital heritage. (October 2002). *Plan for the National Digital Information Infrastructure and Preservation Program, A collaborative Initiative of the Library of Congress*. Retrieved from http://www.digitalpreservation.gov/documents/ndiipp_plan.pdf

Reference Model for an Open Archival Information System (OAIS). (2002). *Recommendation for Space Data Systems Standards, Consultative Committee for Space Data Systems, CCSDS 650.0 –B-1*. Retrieved from http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/Ccsds-650.0-B-1.pdf

Silió, Teresa. (septiembre-octubre 2005). Fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System. *El profesional de la información*, 14(5).

Waugh, Andrew., Wilkinson, Ross., Hills, Brendan., & Dell'oro, Jon. (2000). Preserving Digital Information Forever, Digital Libraries, san Antonio TX. *ACM 1-58113-231-X/00/0006*.