Bibliotecas digitales: seguridad en el modelo de referencia OAIS

Bárbara Muñoz de Solano y Palacios
Juan José Prieto Gutiérrez
&
Luis Fernando Ramos Simón

Session: 216 — *Continuity in the face of digital disasters: Disaster planning and recovery for digital libraries* — Information Technology

Resumen:

Las cuestiones relacionadas con el tratamiento de la conservación de documentos en cualquiera de sus soportes (analógico, digital o mixto) no son sencillas y su consideración implica múltiples decisiones respecto a los factores de riesgo que pueden padecer en el tiempo para garantizar el acceso y la preservación.
Los accidentes generan perdidas de identidad cultural y de conocimiento, siendo por ello necesario disponer de planes de seguridad con el propósito de reducir al máximo la siniestralidad.
Desde el punto de vista del modelo de referencia Open Archival Information System (OAIS), se analizan los tres módulos esenciales que permiten el acceso al documento por parte del usuario: adquisición (Ingest), conservación (Archival storage) y recuperación y acceso a la información, con el objetivo de dar respuesta y solución a variados factores de riesgo inherentes de cada fase.

Palabras clave:
OAIS, Preservación digital, Bibliotecas digitales, Planes de preservación, Migración, Formatos de archivo, Conversión de archivos, Emulación, Encapsulamiento, Estandarización.

Abstract:

Issues related to the conservation treatment of any supporting document (traditional, digital or mixed) aren't simple and their consideration involves multiple decisions about risk factors that may suffer over time in order to guarantee access and preservation.
Accidents generate lost of cultural identity and knowledge it´s still necessary to have security plans in order to minimize accidents.

From the of the reference model Open Archival Information System (OAIS) point of view are analysed the three essential modules that provide access to the document by the user: acquisition (Ingest), preservation (Archival storage) and retrieval and access to information, with the intention of providing technical solutions ao varied risk factors inherent in each stage.

Key words:
OAIS, Digital preservation, Digital libraries, Preservation Planning, Migration, File formats, File conversion, Emulation, Encapsulation, Migration, Standardization.

INTRODUCCION

La mayor parte de los esfuerzos desarrollados para la definición de los factores que aseguran a la seguridad de las unidades de información, generalmente se analizan de forma aislada identificando factores de riesgo puntuales; sin embargo, para diseñar un plan de seguridad global que permita gestionar datos minimizando riesgos en un entorno analógico y digital se debe considerar un modelo de estructura global de unidades de información, así como los factores de riesgo que pueden potencialmente afectar a cada uno de los módulos del sistema. El presente texto proporciona definiciones bien aceptadas de cada uno de los módulos definidos en el Modelo de Referencia OAIS (Reference Model for an Open Archival Information System) y en segundo lugar, revisa los factores de riesgo en la seguridad de la información y, a partir de ellos, sugerir una propuesta de implantación/prevención.

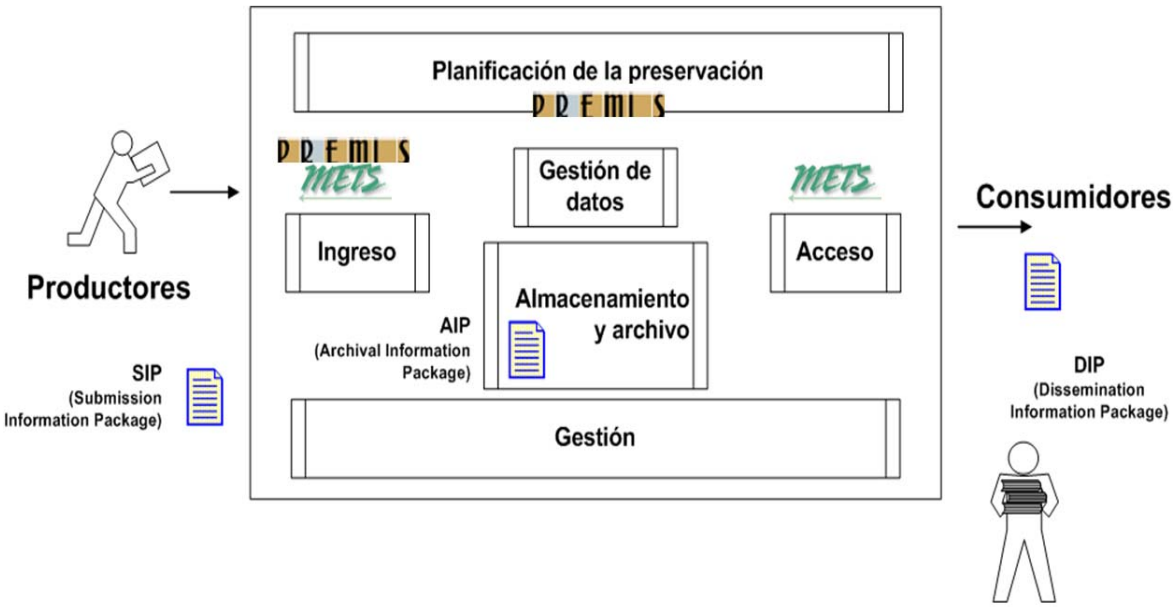

Figura 1: Modelo de referencia Open Archival Information System (OAIS)

El modelo de referencia Open Archival Information System (OAIS) puede considerarse como un punto de partida para cualquier institución que desee implantar un sistema de gestión documental bien sea analógico, digital o mixto ya que OAIS no es una aplicación sino un modelo de referencia que nos va a permitir identificar en cada parte del modelo los aspectos, procesos, tecnologías específicas, estructura de base de datos o plataformas informáticas que deben considerarse en cada módulo para afianzar la seguridad de los datos.

MODELO OAIS

La primera pregunta a la hora de diseñar y planificar la seguridad del sistema de información es la de analizar los posibles riesgos que puede padecer. Este análisis debe responder a preguntas como ¿Qué quiero proteger? ¿Cómo lo puedo hacer? ¿De cuántos recursos económicos dispongo? ¿Qué riesgos puedo padecer?, etc. para todo ello es necesario identificar todos los elementos disponibles en la institución, los tangibles: sistemas de seguridad, ordenadores, CDPs, documentación y los intangibles como los aspectos legales y la propia documentación virtual. Si bien es verdad que el modelo OAIS establece la terminología y los conceptos relevantes para la gestión de flujos y procesos documentales, en el presente artículo únicamente vamos a identificar los principales componentes y procesos para que posteriormente se puedan identificar riesgos en la seguridad de las unidades de información, su prevención y posible subsanación de errores de tal forma que la documentación al tiempo que está disponible para los usuarios finales se asegure su seguridad en tiempo real.

Adquisición (Ingest): los documentos, después de haber sido sometidos a un proceso de selección y valoración previa, durante el proceso de Ingest sufren la primera transformación que se materializa en procesos de verificación del documento. Esta primera fase es vital para asegurar que todo cuanto se introduce en el Sistema de Información no supone un riesgo inicial para la seguridad del mismo. La metodología es distinta según el tipo documental del que se trate, no obstante con el fin de evitar que estos riesgos externos se incorporen durante el Ingest en el Sistema de Información, se hace necesario mantener unos mecanismos de seguridad para afrontar dichos riesgos. Las medidas a aplicar, serian:

Objeto digital

- Validación de la sintaxis y las reglas de integridad en bases de datos
- Formatos de objetos y metadatos que es posible incluir. Este elemento de control es esencial, ya que un formato incorrecto puede desestabilizar los procesos de gestión del sistema (por ejemplo si tiene un tamaño exageradamente grande) o bien por la multiplicidad de formatos que posteriormente dificulten la seguridad de gestión del sistema
- Instalación de antivirus, cortafuegos, detectores de intrusos, antimalware, etc. Las soluciones deben ser capaces de una detección proactiva que permitan bloquear códigos maliciosos e incluso los desconocidos.
- Identificadores persistentes
- Actualización periódica del sistema operativo y de todas las aplicaciones instaladas en los ordenadores. Aumentando el nivel de seguridad y minimizando la posibilidad de ser victimas de usuarios mal intencionados.

- Bloqueo de páginas Web de contenido malicioso. La filtración de direcciones que contengan un contenido no seguro es aconsejable.
- Durante el proceso de ingesta de los documentos en el sistema se añade toda la información descriptiva necesaria. No obstante, para asegurar la preservación digital y evitar o identificar posibles errores/fallos de seguridad es necesario contar con la información adicional que no se incluye en formato analógico, y que se recoge en el conjunto específico de metadatos PREMIS más orientados a asegurar el material. La información que incluye esta tipología de metadatos es por ejemplo el programa utilizado para la creación del documento, la versión del mismo, el propietario, información sobre la licencia con la que se publicará el contenido, etcétera.

Objeto analógico

- Identificar debidamente la procedencia de la fuente de la información sobre el contenido, quién ha tenido su custodia desde su origen y su historia (incluida la historia del procesamiento).
- Analizar la integridad del Documento, que no falten partes del mismo, ni que esté dañado el soporte.
- Control de fiabilidad de los proveedores
- Elaborar y mantener actualizada la lista de usuarios que tengan permisos para aceptar donaciones de material, adquisiciones, etc... con especificación del nivel de responsabilidad que tiene cada uno de ellos.

Conservación (archival storage). Una vez superada la fase de incorporación, los documentos físicos se ubican en los depósitos o estanterías de sala de libre acceso tras haber sido catalogados, mientras que los objetos digitales deben almacenarse en el repositorio tras su asignación de metadatos descriptivos. En términos de seguridad debe considerarse si en lo que respecta a objetos digitales se dota o no de dos dispositivos de almacenamiento diferenciado, uno para el master digital y otro para el derivado, ya que sería más lógico dedicar más recursos de seguridad al depósito de master que de derivados. En documentación digital no se trata de generar una copia exacta de los master para difusión, sino de obtener un documento óptimo para su difusión; una manifestación con características diferentes que faciliten la rápida difusión y acceso al contenido del documento, aunque con ello se pierda perdurabilidad del derivado.
Una vez más las medidas que es necesario considerar en lo que respecta a seguridad son de diferente aplicación dependiendo del tipo de documento del que tratemos:

Objeto digital

- *"Si en el contexto del material impreso, la preservación es básicamente un problema de supervivencia del objeto físico en que se identifica un problema y a raíz de éste se desarrollan las estrategias de preservación. En el contexto digital, la preservación es un proceso continuo, de plazos cortos y*

exige una inversión constante"[1] Por este motivo es necesario llevar a cabo mecanismos para comprobar si hay errores, si se pueden evitar, y que el documento conserva el "*look and feel*"[2] en estructura contenido y presentación:

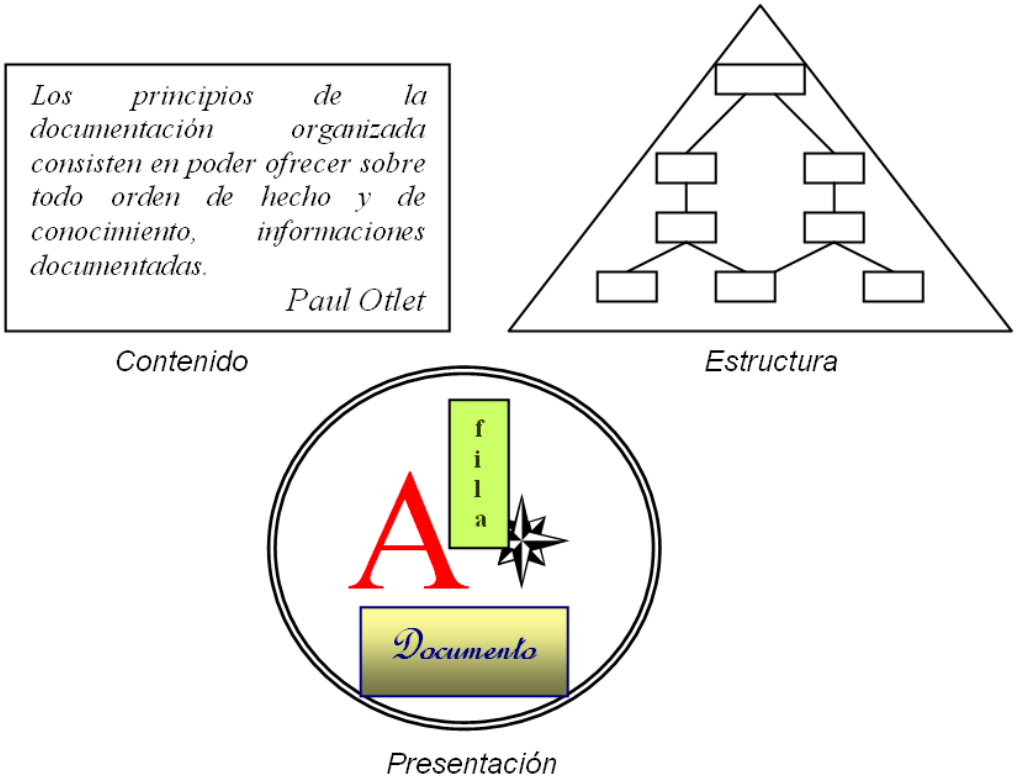

Figura 2: "Look and feel"

Las medidas de prevención que se relatan a continuación tienen la misión de evitar la obsolescencia tecnología de hardware y software asegurando la consulta futura al documento digital:

1. Emulación: la emulación, como su nombre indica, en lugar de conservar físicamente la tecnología original, emula el comportamiento de las viejas plataformas tecnológicas y sus sistemas operativos "*Es el proceso mediante el cual se diseña y se instala un nuevo sistema informático capaz de simular las funciones que realizaba otro sistema ya obsoleto y generalmente de características*

[1] Sanllorenti, A. Mª. "Digitalización y preservación digital". En: *VI Encuentro de Bibliotecas Universitarias: (Buenos Aires, abril de 2001).* Asociación de Bibliotecarios Graduados: [Buenos Aires], 2001; p. 5

[2] Este término inglés se refiere a los diferentes aspectos y funcionamiento del objeto digital. De esta manera, podríamos definir en este contexto la palabra Feel como "la forma en que la interface se comunica con el usuario por sí misma; es decir, los aspectos visuales, el diseño gráfico propiamente dicho, la estética (tipografía colores, etc.)"; mientras que "Look" hace referencia al como la interfaz muestra el contenido, es principalmente la estructura, los menúes, las convenciones de usabilidad que dan sentido al contenido en caso de que éste tenga estructura.

diferentes, con el fin de poder ejecutar los viejos programas informáticos del segundo"[3].

2. Migración: "*Es la transferencia periódica del material digital desde una configuración de hardware o software a otra; o bien desde una generación de tecnología informática a la siguiente"*[4] . En esencia, supone la transformación de la forma lógica de un objeto digital a diferencia de la estrategia de "renovación del soporte", la cual mantiene el flujo de datos transfiriéndolos de un soporte a otro. Dada la actual falta de permanencia de los soportes y su rápida evolución la técnica exige un esfuerzo incesante fundado en cadenas constantes y frecuentes de migraciones[5]. Se estima que la copia de bits a nuevos soportes llegará a ser necesaria cada cuatro años. En lugar de centrarse en conservar los componentes tecnológicos, la estrategia de la migración busca asegurar el contenido informativo de cada documento sabiendo que el proceso de transferencia abarca, en la mayoría de los casos, además de los ficheros seleccionados:

    - *Las reglas de manipulación de los datos y la información sobre la procedencia y el contexto original.*
    - *Los metadatos asociados a cada registro.*
    - *La información sobre los derechos, incluidas las licencias*
    - *Datos sobre las herramientas necesarias para asegurar el acceso*

[3] Granger, Stewart. "Emulation as a digital preservation strategy"[En Línea]. En: D-Lib Magazine. 2000, 6(10):16. Disponible en Web: http://www.dlib.org/dlib/october00/granger/10granger.html [Fecha de consulta: 18 de abril de 2012].

[4] Preserving digital information: report of the Task Force on Archiving of Digital Information [en línea]; commissioned by the Commission on Preservation and Access and the Research Libraries Group. Washington, D.C.: Commission on Preservation and Access, [c1996]. Disponible en Web: http://www.rlg.org/ArchTF/tfadi.index.htm [Fecha de consulta: 10 de mayo de 2012].

[5] El Consultative Committe for Space Data Systems (CCSDS) considera que la migración se subdivide en cuatro categorías:

1) Copia (refreshment): el documento resultante del proceso mantiene una secuencia de bits idéntica a la del original.

2) Réplica (replication): asegura la disponibilidad del objeto digital; si entendemos por disponibilidad la manejabilidad del paquete de información.

3) Reempaquetado (repacking): al igual que lo anterior asegura la disponibilidad del objeto digital.

4) Transformación (transformation): modifica la secuencia de bits del documento original. Es la idea que mejor define, sin duda alguna, el proceso de la migración es esta investigación. Según Paul Wheatley, la migración es un concepto extremadamente amplio. Con el objetivo de aplicar el tratamiento correcto de conservación a cada documento establece la siguiente subdivisión:

1. Minimum migration.
2. Preservation migration.
3. Automatic conversion migration.
4. Recreation human conversion migration.

- Realización de copias de respaldo (backups), de seguridad y recuperación de datos y su periodicidad. Cabe la posibilidad, por parte de la institución, de realizar las copias de seguridad sobre soportes físicos externos. Dichos dispositivos de almacenamiento deben mantenerse bien diferenciados, uno para el master y otro para el derivado, y hospedados en espacios físicos diferentes, a ser posible fuera del edificio, con el fin de asegurar la perdurabilidad de los documentos, para este fin se puede contar con empresas dedicadas a ello.

Objeto analógico

Una de las máximas de la seguridad es que la seguridad total no existe, es imposible la eliminación total del riesgo, sino la reducción. Las medidas empleadas para atajar las incertidumbres deben ser coherentes y acordes al análisis previo realizado por la institución. Las medidas que debe disponer el centro deben ser organizativas, y físicas en lo que respecta a la seguridad del almacenamiento de la colección analógica:

- Es necesaria tanto la colaboración de los trabajadores como la de los usuarios y para ello deben trazarse planes de cooperación en los cuales, se buscará tener constancia de incidencias, recomendaciones, consejos y puntos de vista que puedan suceder en torno al correcto o defectuoso servicio de acceso al documento y en definitiva, poder solucionarlo para ofrecer el servicio requerido por el centro sin riego para la seguridad de los documentos.
- Control de accesos físicos al edificio para garantizar la integridad del patrimonio documental. Las mediadas que deben llevarse a cabo son: presencia del servicio de vigilancia, videovigilancia, lectores de tarjetas para identificación del usuario (biometría, RFID, tarjetas magnéticas), circuito de CCTV, sensores de detección de movimiento, etc.
- Asegurar la autenticidad documental de la información, de forma que el contenido informativo de lo que se almacena en los depósitos no sea modificado. Las medidas de seguridad que deberán ser implantadas como mínimo serán las siguientes:
  - Protocolos de apertura y cierre.
  - Acceso restringido a personal autorizado, mediante claves, llaves, tarjetas electrónicas, sistemas de reconocimiento biométrico, empleo de sistemas autónomos de control de accesos, tecnología de proximidad, etc.
- Actualización periódica del sistema de catalogación y de todas las aplicaciones instaladas en los ordenadores. Aumentando el nivel de seguridad y minimizando la posibilidad de ser victimas de obsolescencia tecnológica que imposibilite recuperar información descriptiva de los documentos y por tanto su identificación física ulterior en los depósitos

<u>Recuperación y Acceso a la información</u>:

Se lleva a cabo en dos etapas, puesto que los documentos generalmente no están en un formato ni en un depósito al que puedan acceder los usuarios. Así que se generan dos fases.

- Búsqueda de la referencia del documento en el catalogo o biblioteca digital generado en el propio sistema a partir de los metadatos.
- Traslado de ese documento desde el depósito accesible para los usuarios durante el tiempo de consulta o conversión del documento a un formato que permita su difusión

Las consecuencias de entradas fraudulentas o ataques a los sistemas informáticos de la institución a partir del módulo de acceso son múltiples, yendo desde la manipulación y robo de la documentación en las propias salas de consulta, hasta la modificación de la Web institucional, eliminación de programas y aplicaciones, ataques de suplantación de identidad, captura de cuentas de usuario y contraseñas, modificación del tráfico y de las tablas de enrutamiento, introducción en el sistema de malware, ataques de inyección de código (SQL) y de denegación de servicio (DoS), ataques contra los sistemas criptográficos, etc. Es aconsejable vigilar e incluso limitar los puertos usb en los dispositivos de acceso público y limitar aquellos utilizados por el personal de la biblioteca. Por atraparte, el empleo de dispositivos móviles, smartphone y tabletas, es cada día más frecuente para acceder a los servicios de las instituciones desde lugares remotos. Esto genera la plena exposición, tanto del usuario como de empleados, a variados tipos de ataques informáticos. Por este motivo, se hace necesario el empleo de soluciones antivirus que respondan a la especificidad de los dispositivos móviles.

Conclusión

El acceso a la información, especialmente relacionada con sistemas basados en el uso de tecnologías de la información y comunicaciones, cada día es más importante, por lo que incidentes y fallos tienen un severo impacto: la deficiencia e incluso la imposibilidad de acceso al patrimonio bibliográfico.
El empleo de las soluciones aportadas en el documento, en las fases de adquisición, conservación y recuperación y acceso a la información permitirán mantener las premisas del modelo OAIS: acceso al documento y preservación del mismo.

La gran importancia del tema de la seguridad justifica la redacción de pautas y consejos a aplicar que, ante cualquier súbita alarma que involucre al documento en cualquiera de sus soportes, con independencia del percance y alcance de la misma, el personal del centro actuará de una manera ordenada siguiendo unas pautas de intervención previamente establecidas para garantizar y preservar la memoria bibliográfica y así poder trasmitirla.

Bibliografía:

- Agenjo, Xavier; Hernández; Francisca. La biblioteca virtual: función y planteamiento. Disponible en: http://eprints.rclis.org/bitstream/10760/14352/1/La_biblioteca_virtual_final.pdf [Consultado el 05 de mayo de 2012]
- Allinson, Julie. OAIS as a reference model for Repositories. Digital Repositories Support, UKOLN, University of Bath, 2006. Disponible en: http://eprints.whiterose.ac.uk/3464/1/Drs-OAIS-evaluation-0.5.pdf [Consultado el 18 de abril de 2012].
- Beedham, Hilary; Missen, Julie; Palmer, Matt; Ruusalepp, Raivo. Assessment of UKDA and TNA compliance with OAIS and mets standards. Disponible en: http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf [Consultado el 10 de mayo de 2012]
- Chowdhury, Gobinda. From digital libraries to digital preservation research: the importance of users and context, Journal of Documentation, 2010, Vol. 66 Issue: 2, pp.207 – 223
- Goodyear, Marilu; Richard Fyffe, Institutional Repositories: An Opportunity for CIO Campus Impact, EDUCAUSE Review, Vol. 41, No. 2, March/April 2006, pp. 10–11. Disponible en: http://www.educause.edu/apps/er/erm06/erm0626.asp [Consultado el 10 de mayo de 2012]
- Granger, Stewart. Emulation as a digital preservation strategy. D-Lib Magazine. 2000, 6(10):16. Disponible en: http://www.dlib.org/dlib/october00/granger/10granger.html [Consultado el 20 de abril de 2012].
- Michael Day. Preservation metadata initiatives: practicality, sustainability, and interoperability. Disponible en: http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf [Consultado el 12 de mayo de 2012]
- Muñoz de Solano Palacios, Bárbara. La importancia de utilizar metadatos PREMIS. Primer paso para la preservación de objetos digitales. v congreso Nacional de Bibliotecas Públicas. Gijón, 3, 4 y 5 de noviembre. 2010. **http://hdl.handle.net/10421/4883**
- Prieto Gutiérrez, Juan José. Seguridad sobre el patrimonio documental en las bibliotecas. Tesis. Universidad Complutense de Madrid. 2011.
- RLG/OCLC. Preservation Metadata and the OAIS Information Model. A Metadata Framework to Support the Preservation of Digital Objects. A Report by the OCLC/RLG Working Group on Preservation Metadata. 2002. Disponible en: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf [Consultado el 10 de mayo de 2012]
- Rusbridge, Chris. Excuse me ... some digital preservation fallacies? Ariadne, issue 46, Feb 2006. Disponible en: http://www.ariadne.ac.uk/issue46/rusbridge/ [Consultado el 12 de mayo de 2012]
- UNESCO. Directrices para la preservación del patrimonio digital. 2003 Disponible en: http://unesdoc.unesco.org/images/0013/001300/130071s.pdf [Consultado el 14 de mayo de 2012]