

Retos y alternativas para la preservación a largo plazo de información digital en bibliotecas

Challenges and alternatives for long-term preservation of digital information in libraries

Ing. Edisnel Carrazana Castro

Resumen: El presente trabajo describe los problemas fundamentales a los que se enfrentan las bibliotecas en cuanto a la preservación a largo plazo de la información digital, dejando en evidencia la necesidad de tomar acciones para proteger el patrimonio digital de las amenazas a las que está expuesto. Se realiza una síntesis de las soluciones tecnológicas más importantes, las cuales constituyen puntos de referencia en cuanto a lo que se está realizando en el campo de los sistemas de preservación digital. Se valora de forma general el estado actual de la preservación digital a largo plazo en bibliotecas y las alternativas de software más importantes.

Palabras clave: Preservación a largo plazo; sistemas de preservación digital; información; OAIS; bibliotecas

Abstract: *This paper describes the fundamental problems that libraries face in terms of long-term preservation of digital information, demonstrating the need to take action to protect the digital heritage against threats to which it is exposed. A summary of the most important technological solutions is performed, which are points of reference in terms of what is being done in the field of digital preservation systems. In general, it is valued the current state of digital long-term preservation in libraries and major software alternatives*

Keywords: *Long-term preservation, digital preservation systems, information, OAIS, libraries*

Introducción

Desde sus orígenes 4000 o 5000 años atrás, las bibliotecas han tenido un rol fundamental en la sociedad. Desde un inicio marcado por el énfasis en coleccionar, organizar y preservar documentos contables y administrativos en los antiguos territorios mesopotámicos, hasta convertirse en un instrumento auxiliar de la enseñanza y la investigación (Linares, 2004). Las bibliotecas pasaron de

Ing. Edisnel Carrazana Castro: Profesor de Inteligencia Artificial. Departamento de programación. Facultad 2. Universidad de las Ciencias Informáticas. La Habana, Cuba.
ecarrazana@uci.cu

ser meros depósitos de documentos a organizaciones que facilitan la gestión de la información y el conocimiento.

Preservar la información y el conocimiento a través del tiempo y el espacio, ha sido una preocupación en mayor o menor medida de acuerdo a la época y el contexto. La escritura como modo de representar la información más comúnmente usado, ha adoptado diversos soportes desde la antigüedad hasta la época actual, teniéndose tabletas de arcilla, papiro, pergamino, papel y con mayor proporción hoy, documentos digitales.

El papel ha sido el soporte físico de los fondos y servicios de una biblioteca tradicional. De esta variante en papel compuesta por libros, revistas, catálogos, boletines informativos, etc., ha evolucionado en pocos años hacia la biblioteca electrónica, en la que los fondos tienen presencia física de los documentos, pero los servicios se automatizan paulatinamente. En un espacio de tiempo aún más corto, asistimos a una nueva realidad; las bibliotecas usan ya la tecnología disponible y los servicios en red, por lo que, no sólo los servicios, sino la misma información se crea, se procesa y se difunde en forma electrónica, a través de las redes de transmisión de datos (Pérez, Lara, & Naranjo, 2004).

El desarrollo y uso masivo de los medios y tecnologías digitales de procesamiento, transmisión y almacenamiento han traído consigo la disponibilidad de una considerable cantidad de información, debido a una mayor eficacia en los procesos de creación, organización y difusión de la información. En el entorno digital, donde la información se registra, almacena y difunde de forma diferente a lo tradicional, el acceso y la preservación plantean nuevos retos para las bibliotecas.

A pesar de las ventajas que ofrece el entorno digital, a la vez impone serias amenazas en cuanto a la preservación de documentos digitales a largo plazo. No solo están presentes las amenazas asociadas a la degradación de los medios de almacenamiento digitales como el CD-ROM o discos duros, el borrado accidental de la información por parte de usuarios y operadores, o accesos no autorizados con el fin de borrar o alterar la información; también hay que tener en cuenta el ciclo cada vez más rápido de obsolescencia de los equipos de informática (hardware), del software y de los formatos.

Ante tales amenazas de pérdida de información y teniendo en cuenta que muchos servicios en las bibliotecas se prestan en ambiente digital, la presente contribución se plantea como problema:

¿Cómo proteger el patrimonio digital de las bibliotecas de las amenazas a las que está expuesto? La presente investigación tiene como objetivo fundamental describir las alternativas de software más importantes para la preservación a largo plazo de información digital en bibliotecas. La descripción de estas alternativas puede servir como punto de partida para un estudio más profundo de las herramientas de software a elegir para implementar la preservación a largo plazo en las bibliotecas.

El problema de la preservación digital a largo plazo

La preservación digital puede definirse como el conjunto de los procesos destinados a garantizar la continuidad de los elementos del patrimonio digital durante todo el tiempo que se consideren necesarios (Biblioteca Nacional de Australia, 2003).

El mayor impedimento técnico para la preservación a largo plazo es la incompatibilidad entre el hardware y el software utilizado originalmente para crear, almacenar, gestionar y difundir la información digital y las plataformas informáticas actuales o futuras. La falta de interoperabilidad entre las generaciones más antiguas y más recientes de plataformas informáticas es evidente cada vez que los medios de almacenamiento, formatos de almacenamiento de datos, aplicaciones, lenguajes de programación, esquemas de representación, o herramientas de renderizado son reemplazados por nuevos métodos sin compatibilidad hacia atrás (Hedstrom, 2001).

De modo que no basta con preservar la secuencia de bits que constituyen cualquier documento digital (dígase imágenes, videos, audio, código fuente, documentos de texto, bases de datos, animaciones, etc) sino que es necesario garantizar que este permanezca accesible y pueda ser leído en los dispositivos y aplicaciones informáticas disponibles en el momento en que se quiera usar.

Otro concepto que es fundamental en cuanto a preservación digital es el de autenticidad, que es la cualidad de un documento de "...ser lo que pretende ser, inmune a cualquier tipo de manipulación y corrupción" (Duranti, 2001). En otras palabras la autenticidad es el mantenimiento en el tiempo y el espacio de la confiabilidad del documento (Mena, 2006). Pese a que los documentos digitales preservados en bibliotecas no necesariamente constituyan evidencias de hechos o actos de las administraciones, sí tienen valor histórico e intelectual, por lo que debe lograrse que los mismos

permanezcan libres de manipulaciones o corrupciones y sean confiables para los usuarios que los consultan.

A pesar de la importancia de la preservación digital, el papel fundamental de las bibliotecas y archivos digitales para asegurar el futuro acceso a la información con valor permanente ha pasado a segundo plano para mejorar el acceso a los materiales actuales y utilizados activamente. Como consecuencia de ello, la preservación digital se mantiene en gran parte experimental y repleta de los riesgos asociados con los métodos no probados, y los requisitos de preservación digital no se han tenido en cuenta en la arquitectura, la asignación de recursos, o la planificación de las bibliotecas digitales (Hedstrom, 1998).

También hay que considerar otras cuestiones que hacen más complejo el problema, y no solo las de tipo técnicas, sino también legales, pues si no se tiene el permiso del creador de un documento, no podrá ser preservado; de índole económica, ya que si no se dedican los recursos adecuados no se podrá preservar el patrimonio digital a lo largo de los años; y también de tipo institucional, pues es necesario que en las instituciones exista el compromiso de todos los involucrados, así como el establecimiento de planes y procedimientos para lograr la preservación digital a largo plazo.

La necesidad de preservar la información digital en las bibliotecas

La gran cantidad de información digital presente hoy en muchas bibliotecas se debe, por una parte, a proyectos de digitalización de materiales raros y valiosos, que generan continuamente un gran volumen de documentos digitales. Por otra parte, muchos de los materiales provenientes de los más diversos proveedores, son adquiridos por las bibliotecas en formato digital, habiendo sido creados en este entorno.

Son muchas ya las instituciones que están dedicando esfuerzos y recursos a la investigación y desarrollo de sistemas para la preservación de información digital, entre ellas la UNESCO, que ha impulsado iniciativas como la Carta para la preservación del Patrimonio digital, y las

Directrices para la preservación del patrimonio digital.

La carta es un documento no vinculante desde el punto de vista jurídico, a través del cual la

UNESCO alerta a los Estados miembros del peligro de pérdida que acecha al patrimonio digital, instándoles a tomar medidas jurídicas e institucionales para protegerlo. El documento titulado *Directrices para la preservación del patrimonio digital*, fue preparado por la Biblioteca Nacional de Australia, a instancias de la UNESCO, y publicado en 2003 (Orera, 2008).

En la Carta, en el artículo 4 se expresa "A menos que se haga frente a los peligros actuales, el patrimonio digital desaparecerá rápida e ineluctablemente". También en el artículo 9: "Hay que preservar y poner a disposición de cualquier persona el patrimonio digital de todas las regiones, naciones y comunidades a fin de propiciar, con el tiempo, una representación de todos los pueblos, naciones, culturas e idiomas." (Unesco, 2003).

Este llamado de la UNESCO a preservar el patrimonio digital debe involucrar no solo a los estados miembros, sino a todas las instituciones que de una forma u otra pueden ayudar a concebir y establecer el contexto adecuado para la preservación digital. En este escenario, las bibliotecas, ya sean nacionales, universitarias, especializadas o de cualquier tipo, tienen un papel relevante al contar con importantes recursos de información, profesionales altamente preparados y por su responsabilidad ante la sociedad.

Sistemas de preservación digital

Aunque hoy no se puede afirmar que hay una solución definitiva para la preservación digital a largo plazo, si se pueden notar avances en este sentido, tales como el desarrollo de estándares y directrices para las puesta en marcha de planes de preservación digital; la concepción y aplicación de estrategias tecnológicas diversas que pueden complementarse; y el desarrollo de sistemas informáticos que implementan estos estándares y estrategias.

Sin embargo, en la mayoría de las bibliotecas, la preservación no es una finalidad en si misma, sino simplemente una actividad necesaria que ha de permitir dar cumplimiento a su función primordial que es dar acceso a la información y la documentación. Como consecuencia, los procedimientos y las aplicaciones informáticas de preservación se diseñan como apéndices a otros procedimientos o aplicaciones, o como simples módulos o rutinas de aquellos; un ejemplo es el complicado encaje de la preservación con software de repositorios tan usados como DSpace y Fedora (Térmens, 2009).

El objetivo de los sistemas de preservación digital es que la información que contienen siga siendo accesible a los usuarios durante un largo período de tiempo. El problema clave en el diseño de tales sistemas es que el período de tiempo es muy largo, mucho más largo que el tiempo de vida de los medios de almacenamiento individual, componentes de hardware y de software, y los formatos en los que la información se codifica (Rosenthal & Robertson, 2005). Para enfrentar estas complejidades se han desarrollado modelos teóricos, entre los más significativos esta OAIS (Open Archival Information System) y que ha dado lugar a la norma ISO 14721:2003.

Este modelo de referencia se ocupa de una amplia gama de funciones de preservación de la información de archivo, incluyendo la ingesta, almacenamiento de archivos, gestión de datos, el acceso y la difusión. También se ocupa de la migración de la información digital a los nuevos medios y formas, los modelos de datos que se utilizan para representar la información, el papel del software en la preservación de la información y el intercambio de información digital entre los archivos. Es aplicable a cualquier organización con la responsabilidad de hacer que la información este disponible por largo plazo (International Organization for Standardization, 2012).

Los sistemas de preservación digital más importantes en la actualidad tratan de cumplir las pautas descritas en el estándar antes mencionado, y se puede observar que varían en el conjunto de amenazas que consideran importantes, en las estrategias que seleccionan y en la forma de implementarlas (Rosenthal & Robertson, 2005). Algunos de estos sistemas se describen a continuación:

LOCKSS (Lots Of Copies Keep Stuff Safe)

Es una herramienta diseñada para las bibliotecas para asegurar la continuidad en el acceso de la comunidad a las revistas científicas publicadas en la web. Permite a las bibliotecas tomar la custodia del material al que se suscriban, de la misma manera que lo hacen para el papel, y preservarlo. Al preservarlo se aseguran de que, para su comunidad, los vínculos y las búsquedas se mantengan para obtener el material publicado, incluso si ya no está disponible en la editorial. LOCKSS permite a las bibliotecas recoger contenido a medida que se publica y cooperar en una red par a par para detectar y reparar documentos dañados o ausentes. Se puede ejecutar en hardware de computadoras genéricas usando software de código abierto y casi

no requiere administración especializada, haciendo el costo de la preservación manejable (Reich & Rosenthal, 2001).

Pórtico

Es un servicio de preservación digital de revistas electrónicas, libros y otros contenidos, de ITHAKA, una organización sin fines de lucro dedicada a ayudar a la comunidad académica a utilizar las tecnologías digitales para preservar el registro académico y para avanzar en la investigación y la docencia de forma sostenible. Portico entiende la preservación digital como la serie de políticas y las actividades de gestión necesarias para garantizar la usabilidad, la autenticidad, la capacidad de descubrimiento y la accesibilidad de los contenidos en largo plazo. Sirve como un archivo permanente de los contenidos de más de 117 editores, en nombre de más de 2.000 sociedades científicas y asociaciones (Morrissey, Meyer, & Bhattarai, 2010).

DIAS

IBM diseñó DIAS (Digital Information Archiving System) para la Biblioteca Nacional de los Países Bajos y se convirtió en el núcleo de e-Depot, su solución de archivo digital a largo plazo. DIAS reclama ser un sistema que cumple el estándar OAIS, pero en el corazón parece ser una vieja base de datos DB2 con una interfaz para almacenamiento. Es típico de una solución de proveedor que complace las preocupaciones de los bibliotecarios sin abordar realmente los problemas clave de ingeniería y ciencias de la computación. Dentro del proyecto KOPAL, IBM mejoró DIAS para permitir el uso cooperativo y dar soporte a la migración de formatos de archivo (Seadle, 2013).

e-Depot

e-Depot entró en funcionamiento en enero de 2003 y fue específicamente diseñado para almacenar y mantener los objetos digitales de manera indefinida, de acuerdo al mandato de la Biblioteca Nacional de los Países Bajos como depósito nacional. (Ras, 2009). E-Depot es un ambiente de archivo digital para el acceso permanente a fuentes de información digital. Su objetivo es almacenar el depósito holandés de publicaciones electrónicas y asegurar el acceso permanente a información científica para la comunidad de investigadores. El núcleo de e-Depot es DIAS, basado en el modelo de referencia OAIS, un estándar internacional. E-Depot ha sido desarrollado para facilitar el almacenamiento permanente y el acceso a largo plazo;

procedimientos de migración y emulación están siendo desarrollados como estrategias para el acceso permanente (Vernooy, 2009).

KOPAL

El sistema KOPAL trata con el problema de la obsolescencia de formatos de archivo a través de la migración de los formatos a lo largo de su arquitectura. Cada paquete de archivos está en un formato llamado Formato de Objeto Universal, que describe una estructura para almacenar los metadatos de preservación junto con los archivos de contenido. Los archivos de contenido pueden ser de cualquier formato. Cada formato de archivo se hará obsoleto en algún punto en el tiempo, la idea es identificar los archivos en peligro antes de que suceda. Dentro de este sistema pueden ser manejadas varias versiones migradas de objetos y metadatos adicionales pueden describir cada proceso de migración (Seadle, 2013).

LuKII

Específicamente este proyecto propone la interoperabilidad entre los elementos open-source de LOCKSS Y KOPAL, con el fin de combinar la preservación de flujos de bits de forma rentable con una herramienta establecida para la facilidad de mantenimiento y migración de formato. Basado en estas metas, los elementos principales de este proyecto son: establecer una red de sistemas LOCKSS rentable en Alemania incluyendo la infraestructura para proveer apoyo técnico constante y administración para LOCKSS y sus variantes; conceptualizar e implementar la interoperabilidad entre LOCKSS y KOPAL; probar el prototipo de interoperabilidad archivando los datos provenientes de los repositorios institucionales alemanes (Seadle, 2013).

PANDORA

La Biblioteca Nacional de Australia, posee una amplia experiencia en el desarrollo de metodologías para recoger y archivar publicaciones web, ya que desde 1996 recoge y gestiona sitios web australianos a través de la base de datos PANDORA. Además, ha desarrollado un sistema propio para recoger y archivar recursos web, denominado PANDAS (Pandora Digital Archiving System). La biblioteca mantiene el portal PADI (Preserving Access to Digital Information) que recoge abundante información sobre el tema de preservación digital, incluyendo resúmenes periódicos sobre los principales avances en este campo (Orera, 2008).

Los sistemas descritos anteriormente de forma general y cumpliendo con las pautas del modelo de referencia OAIS, implementan las estrategias más usadas actualmente para lograr la preservación digital a largo plazo, estas son (Biblioteca Nacional de Australia, 2003):

- Migración: Básicamente consiste en transferir materiales digitales de una generación de un equipo o un programa informático a otra. Distinta del refrescamiento, que mantiene el flujo de datos transfiriéndolos de un soporte a otro, la migración supone la transformación de la forma lógica de un objeto digital de modo que el objeto conceptual pueda ser restituido o presentado por nuevos equipos o programas informáticos
- Emulación: La emulación consiste en utilizar programas informáticos que hacen funcionar una tecnología con las características de otra. En el contexto de la preservación digital a largo plazo, esta estrategia podría hacer que las futuras tecnologías se comportaran como el entorno original de un objeto digital preservado, de modo que éste podría presentarse en su forma original, a partir del flujo de datos original.

Conclusiones

Del estudio realizado y sin ánimo de establecer comparaciones entre sistemas, se puede decir que los intentos más serios en cuanto a dar solución a los problemas de la preservación son: KB e-Depot de la Biblioteca Nacional de los Países Bajos, KOPAL de la Biblioteca Nacional de Alemania, LOCKSS de la Universidad de Stanford y Portico de Ithaka. Estos incorporan supuestos diferentes acerca de cómo tratar la preservación a largo plazo, reciben importante apoyo financiero y el reconocimiento de sus gobiernos nacionales. Aunque no tienen el mismo modelo de negocios, cuentan con el respaldo de numerosas comunidades de investigadores y desarrolladores que garantizan su continuidad. Resaltar el hecho de que LOCKSS es software de código abierto y está disponible su descarga en internet, con lo que es más fácil de adquirir que las demás alternativas analizadas.

En el ámbito bibliotecario es una necesidad el establecimiento de políticas de preservación a largo plazo, sin pasar por alto en las mismas la importancia de conservar las cualidades de confiabilidad y autenticidad de la documentación digital.

Especial atención debe darse a los requisitos funcionales del software de gestión que se utiliza para implementar bibliotecas digitales, pues de forma general tienen debilidades estructurales que no les permiten preservar a largo plazo la información digital.

La preservación digital a largo plazo se enfrenta a problemas todavía no resueltos, utilizando estrategias que en la actualidad están

siendo probadas y que requieren de más investigación. La descripción de cada una de las alternativas brinda elementos a tener en cuenta para realizar una adecuada selección de herramientas de software para la preservación digital a largo plazo. ■

Recibido: mayo de 2014

Aceptado: julio de 2014

Bibliografía

- Biblioteca Nacional de Australia. (2003). *Directrices para la preservación del patrimonio digital*. Recuperado a partir de <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>
- Duranti, L. (2001). Concepts, Principles, and Methods for the Management of Electronic Records. *The Information Society*, (17), 271-279.
- Hedstrom, M. (1998). Digital preservation: a time bomb for Digital Libraries. *Computers and the Humanities*, (31), 189-202.
- Hedstrom, M. (2001). *Exploring the Concept of Temporal Interoperability as a Framework for Digital Preservation*. Recuperado a partir de <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.22.1426>
- International Organization for Standardization. (2012). *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*. Recuperado a partir de http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=57284
- Linares, R. (2004). La Bibliotecología y sus orígenes. *Ciencias de la Información*, 35(3). Recuperado a partir de cinfo.idict.cu/cinfo/article/view/134
- Mena, M. M. (2006). *Propuesta de requisitos funcionales para la gestión de documentos archivísticos electrónicos en la administración central del estado cubano* (Tesis para optar por el grado de Doctor en Ciencias de la Información). Universidad de La Habana, Ciudad de La Habana.
- Morrissey, S. M., Meyer, J., & Bhattarai, S. (2010). *Portico: A Case Study in the Use of the Journal Archiving and Interchange Tag Set for the Long Term Preservation of Scholarly Journals*. Recuperado a partir de <http://www.ncbi.nlm.nih.gov/books/NBK47087/>
- Orera, L. (2008). *Preservación digital y bibliotecas: un nuevo escenario*. Recuperado a partir de <http://hdl.handle.net/10760/15142>
- Pérez, M. S., Lara, Y., & Naranjo, D. (2004). *La Digitalización de textos, Base de la Biblioteca Digital; su impacto como apoyo a la Docencia en la UCI*. Recuperado a partir de http://www.dtic.co.cu/FTP/libros/digitaliz_textos.pdf
- Ras, M. (2009). The KB e-Depot: Building and Managing a Safe Place for e-Journals. *Liber Quarterly*, 19(1). Recuperado a partir de <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-113555/8232>
- Reich, V., & Rosenthal, D. S. (2001). LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, 7(6). Recuperado a partir de <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/june01/reich/06reich.html>
- Rosenthal, D., Robertson, T., Lipkis, T., Reich, V., & Morabito, S. (2005). Requirements for Digital Preservation Systems: A Bottom-Up Approach. *D-Lib Magazine*, 11(11). Recuperado a partir de <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
- Seadle, M. (2013). EU digital preservation projects. En *Digital Archiving* (pp. 47-50). Humbolt-Universität zu Berlin.
- Térmens, M. (2009). *Los archivos y las bibliotecas ante la preservación digital: ¿un solo enfoque?* Recuperado a partir de <http://hdl.handle.net/10760/13883>
- Unesco. (2003). *Carta sobre la preservación del patrimonio digital*. Recuperado 24 de enero de 2014, a partir de http://portal.unesco.org/es/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html
- Vernooy, M. (2009). *Enhanced Publications: Linking Publications and Research Data in Digital Repositories*. Amsterdam: Amsterdam University Press. Recuperado a partir de books.google.com