

Los metadatos, una herramienta para preservar al conocimiento digital

ALEJANDRO JIMÉNEZ LEÓN
*Dirección General de Servicios de Cómputo
Académico, UNAM, México*

INTRODUCCIÓN

Dividamos el análisis en dos grandes apartados. El primero concerniente a la preservación de información digital y el segundo al uso de metadatos para cumplir el cometido de preservar al contenido digital.

Si observamos el entorno informático evoluciona aceleradamente todos los días, de hecho la vida útil de la información digitalizada es de seis a siete años; el motivo: envejecimiento de los soportes digitales (hardware) y obsolescencia de los formatos (software), así como la incompatibilidad con los nuevos sistemas. Esta situación se debe en parte a los sistemas operativos y las aplicaciones de Microsoft ya que su vida operativa es de seis años en el mercado, lo cual afecta drásticamente el medio informático en tanto que el 90% del mercado de CP utilizan Windows como plataforma principal.

Por tanto es necesario establecer una política de preservación digital que permita:

- ❖ Definir políticas y criterios para determinar qué material debe conservarse.

III Encuentro de Catalogación y Metadatos

- ❖ Diversificar las plataformas en que pueda trabajar ese contenido.
- ❖ Establecer mecanismos que permitan darle continuidad al contenido.
- ❖ Evitar la pérdida de información y garantizar el acceso.

Estas acciones permitirán construir, adquirir y renovar los conocimientos grabados para garantizar la continuidad del ciclo de la información digital, desde la creación hasta su difusión.

DESARROLLANDO CONTENIDO DIGITAL

Al elaborar recursos informativos digitales se fomenta la difusión del conocimiento y que éste llegue a un número cada vez mayor de usuarios, además de favorecer el desarrollo de nuevos productos informativos. De allí que la digitalización de información deba convertirse en una forma de socializar la riqueza cultural de un país a través de las TIC (Tecnologías de la Información y la Comunicación).

Ahora bien todo desarrollo de contenido digital debe tener como premisa el hecho de poder ser usado en años posteriores a su creación, por lo que su preservación es una necesidad imperante para lograr este fin; el desarrollador debe estar consciente de esta necesidad y de cómo la tecnología que hoy favorece al proyecto, el día de mañana pueden implicar el fin del recurso digital. Cabe señalar que el proceso de digitalización por sí solo no implica preservar el conocimiento, para llegar a este punto es necesario establecer una política de preservación digital al momento de procesar el material con el fin de realmente estar preservándolo.

Esto nos lleva a mantener en el largo plazo dos tipos de valores de los documentos:

1. El valor del contenido, texto ilustraciones, etcétera.
2. El valor de la evidencia, los documentos de archivo pueden ser requeridos para probar la existencia de determinadas transacciones, lo cual obliga no sólo a preservar el contenido del documento, sino a certificarlo adecuadamente.

Por lo tanto podemos distinguir tres tipos de aplicaciones de las TIC en la preservación.

1. Proteger los originales, por medio de la creación de copias digitales con la calidad suficiente para que puedan ser utilizadas en lugar de los originales.
2. Representar los originales supone una digitalización de calidad intermedia que permita disponer de un modo integral de la información que se tenía con los originales.
3. Trascender los originales con fines que hubieran sido imposibles de lograr con las fuentes originales.

Si bien es cierto que existe un importante número de proyectos de digitalización, éstos adolecen en muchas ocasiones de falta de lineamientos que los hagan compatibles con otros proyectos nacionales o internacionales. Como ejemplo tenemos el proyecto de *Google*, la compañía de búsquedas por Internet más grande del mundo, que diariamente digitaliza miles de libros para el proyecto books.google.com. Internet Archive que desea crear una moderna Biblioteca de Alejandría que contenga todos los textos y videos de dominio público. *Amazon* ha estado escaneando libros, al igual que *Microsoft* y *Yahoo!*

DEFINICIÓN DE UNA POLÍTICA DE PRESERVACIÓN DIGITAL

Hasta el momento la evolución tecnológica que nos inunda ha provocado que las instituciones no hayan elaborado estrategias de conservación acordes con las necesidades del entorno. Estas políticas deberían controlar desde la selección, creación, vida de la tecnología usada (hardware y software) posibles aplicaciones, políticas de redundancia, copias de seguridad, planificación contra desastres y mantenimiento, y dar pie a la generación de recursos digitales estables y confiables. Cabe señalar que la experiencia a este respecto es escasa, el marco jurídico inestable y los recursos escasean, con el resultado de que los esfuerzos de preservación son inciertos.

CRITERIOS BÁSICOS EN UNA POLÍTICA DE PRESERVACIÓN DIGITAL

- ❖ Estandarizar los medios de gestión documental para garantizar la compatibilidad de la información en las diferentes unidades de la institución.
- ❖ Mantener un programa de actualización permanente de los medios de gestión de documentos digitales para que se garantice la accesibilidad.
- ❖ Capacitar al personal en la migración de documentos a nuevos soportes y tener por escrito los procedimientos.
- ❖ Desarrollar planes de mantenimiento preventivo, correctivo y de seguridad electrónica de los medios de gestión, para garantizar su buen funcionamiento.
- ❖ Establecer políticas y procedimientos de respaldo periódico.
- ❖ Implementar políticas y procedimientos de seguridad para poder acceder a la información.
- ❖ Poseer instalaciones físicas seguras en donde se gestione y conserve la información.
- ❖ Las políticas deben ser revisadas periódicamente.
- ❖ Realizar auditorías informáticas, en las que se verifique el cumplimiento de los procedimientos, estándares y disposiciones vigentes.
- ❖ Homogeneizar el uso de sistemas informáticos para el proceso de gestión documental, con el fin de asegurar la compatibilidad de la información.
- ❖ Cada área debe tener los procedimientos y controles que garanticen el almacenamiento y acceso de la información.
- ❖ Todos los sistemas deben contemplar una bitácora que permita el registro de cada proceso realizado.
- ❖ Los archivos donde se almacene la información deben estar encriptados y ser de acceso restringido.
- ❖ Se debe contar con especificaciones técnicas sobre los programas computacionales de la institución, las cuales se deben ajustar a sus modelos, y deben contemplar por lo menos los siguientes aspectos: modelo conceptual, modelo físico, diccionario de datos, lenguaje de programación, versión utilizada, códigos fuente actualizados y manuales de usuario.

LONGEVIDAD DEL CONTENIDO Y DEFINICIÓN DEL FORMATO DIGITAL

Todo recurso digital debe tener un periodo de vida en cuanto a su contenido dando lugar a material longevo en mayor o menor medida, durante este tiempo se establecerá un programa de mantenimiento que permita enfrentar los cambios tecnológicos sin temor a una pérdida.

Esta acción cubre tres aspectos:

1. Actualización del contenido.
2. Actualización del formato digital.
3. Actualización de ambos.

Deben agregarse acciones de divulgación y promoción para crear una cultura que formule políticas que tomen en cuenta las circunstancias locales, los medios disponibles y las previsiones para el futuro así como la urgencia de llevarlas a cabo.

En cuanto a **la selección del formato digital**, debe considerarse a este factor como la columna vertebral del desarrollo ya que permitirá la renovación periódica y evitará problemas.¹

El primer criterio para definir al tipo de formato digital es que sea ampliamente usado, un estándar a nivel mundial y que sea de preferencia un formato abierto,² –debe evitarse el uso de formatos propietarios–³ lo cual aumenta la posibilidad de que cuando un formato se vuelva obsoleto existan programas para su conversión, como es el caso de ODF, HTML, XML, PDF TIFF, WAV los últimos dos de los cuales son formatos propietarios que son ampliamente usados. Además de los procesos de conversión es necesario validarlos para asegurarse que los archivos producidos cumplen con las normas prestablecidas.

1 El SP3 de Office 2003 provocó incompatibilidad con las anteriores versiones de Word, Excel, Quattro.

2 Como ejemplo estaría ODF, ODT, ODS, ODG y ODP, TXT, HTML, RTF, DjVu, PDF, JPEG, PNG, SVG, Dirac, Ogg., Vorbis, Speex, Theora, FLAC, los cuales no le imponen restricciones al usuario.

3 Formato de archivo protegido por una patente o derechos de autor, y se pueden distinguir por las siguientes extensiones DOC, DWG, MP3, MPEG, TIFF, entre otros.

Es por eso que en muchas ocasiones el formato PDF⁴ es el más usado, además de que permite agregarle determinados candados de seguridad.

El riesgo de usar un formato propietario es que si la firma que posee el derecho del formato deja de hacerlo, entonces todos los que utilizaron ese formato en el pasado pueden perder la información de sus archivos.

Ahora bien el cambio de formato digital puede ser realizado a través de la **migración** en caso de no existir una nueva versión del formato usado. En la mayoría de las ocasiones los formatos cambian cada dos años o menos y provocan estos problemas. Por ello es recomendable monitorear en qué momento será conveniente realizar la migración para convertir la información a un nuevo formato, de lo contrario el contenido caerá en la obsolescencia.

El proceso de migración tiene dos desventajas:

Es una tarea pesada y los datos originales al ser modificados corren el riesgo de corromperse y dejar de funcionar.

Otra alternativa es la **preservación de los sistemas digitales**, la cual consiste en mantener al sistema en funcionamiento para preservar a los recursos digitales con el sistema en que fueron creados, almacenados y consultados. El problema es que esta medida con el transcurso del tiempo se volverá difícil, pues los repuestos de hardware y software necesarios para mantener y reparar al equipo dejarán de existir y se llegará tarde o temprano a la obsolescencia. Por otra parte debe asegurarse que **los recursos digitales trabajen en diversas plataformas** como Linux, OS/2, Windows o MS-DOS, con esta acción se habrán creado materiales a los que pueden accederse desde diferentes sistemas operativos.

Los criterios antes citados deben contemplarse en el diseño de una política de preservación digital, dejarlos a un lado implicaría que

⁴ Para la preservación de documentos PDF es recomendable el empleo de un UVC (Universal Virtual Computer), otra norma que ayuda a mantener una política de preservación es la ISO 15489 (cuyo objetivo es normalizar las prácticas internacionales en materia de gestión documental, utilizando como punto de partida al estándar australiano AS 4390).

la misma tecnología desechara en el corto plazo esa información. Por lo tanto una política de preservación digital ahorra recursos financieros, materiales y operativos, ya que sus acciones se concentran en el mantenimiento de la obra; sin embargo el factor más importante es evitar la pérdida del conocimiento en el mediano y largo plazo.

LA POLÍTICA DE PRESERVACIÓN DIGITAL ES UNA RESPONSABILIDAD MULTIPARTIDISTA

El primer responsable de ejecutarla debe ser el **Estado** a través de sus instituciones culturales como son las Secretarías de Estado, los Ministerios de Cultura, las Bibliotecas Nacionales, etcétera; y a éstos debe seguir la **Institución** que patrocina el desarrollo de Universidades, Centros de investigación, Bibliotecas y Empresas todos los cuales deben hacer énfasis en que todo proyecto de desarrollo digital cuente con una política de preservación digital que asegure la preservación de los desarrollos institucionales. Quien mayor peso tenga, en la responsabilidad de ejecutar las acciones de preservación será quien se encargue de la elaboración del material, y si obvia estos criterios sin que importe el entorno, entonces los recursos tendrán una vida útil relativamente corta de no más allá de seis años. El **Usuario**, deberá estar atento a los cambios tecnológicos que puede sufrir su información y por lo tanto realizar las modificaciones pertinentes, en el caso de usuarios institucionales será necesario suscribir convenios con sus proveedores de información a fin de asegurarse un respaldo, sin infringir aspectos relativos a la propiedad intelectual del material.

En cuanto a la estructura legal que usa la información digital es diferente a la de otros medios. De manera creciente los objetos de información digital no constituyen una “propiedad” de los usuarios o de las bibliotecas, sólo suponen una autorización de sus creadores, de modo que su uso se rige en términos contractuales.

Por tanto los aspectos legales de la digitalización deben incluir:

- ❖ derechos intelectuales y propiedad del recurso o del software;
- ❖ términos contractuales unidos a un recurso o el hardware y software necesario para acceder a él;

III Encuentro de Catalogación y Metadatos

- ❖ protección de la confidencialidad de individuos e instituciones, y
- ❖ obligación legal de seleccionar y preservar la autenticidad y el contenido de categorías de registros o de recursos individuales.

Aquí es necesario destacar lo que dice la Ley Federal del Derecho de autor⁵ de México, título IV de la protección al derecho de autor, Capítulo IV de los programas de computación y las base de datos.

artículo 105.-El usuario legítimo de un programa de computación podrá realizar el número de copias que le autorice la licencia concedida por el titular de los derechos de autor, o una sola copia de dicho programa siempre y cuando:

Sea indispensable para la utilización del programa, o sea destinada exclusivamente como resguardo para sustituir la copia legítima adquirida cuando ésta no pueda utilizarse por daño o pérdida. La copia de respaldo deberá ser destruida cuando cese el derecho del usuario para utilizar el programa de computación

artículo 110.-El titular del derecho patrimonial sobre una base de datos tendrá el derecho exclusivo, respecto de la forma de expresión de la estructura de dicha base, de autorizar o prohibir:

Su reproducción permanente o temporal, total o parcial, por cualquier medio y de cualquier forma;

Su traducción, adaptación, reodernación y cualquier otra modificación;

La distribución del original o copias de la base de datos

La comunicación al público, y

La reproducción, distribución o comunicación pública de los resultados de las operaciones mencionadas en la fracción II del presente artículo

Por lo anterior existe un problema de preservación en cuanto a las bases de datos en línea, pareciera que al final las instituciones educativas y empresariales las alquilan y su acceso es a través de una licencia, con lo que pasamos de una propiedad física a una virtual, situación que provoca que se pierda el control sobre las colecciones digitales dado que un considerable volumen de información sólo está disponible en formato digital y no existe una versión en papel que pueda ser preservada, por lo que una vez vencida la licencia deja de existir acceso a esos datos. El problema de la preservación digital es cómo detectar el

5 http://sep.gob.mx/wb2/sep/sep_Ley_Federal_del_Derecho_de_Autor#4
"Ley Federal del Derecho de autor [en línea], consultado el 1º de septiembre 2007.

momento en que debemos realizar la transición de hardware o software y es allí donde los metadatos resultan ser una excelente alternativa para realizar este proceso.

DEFINICIÓN DE METADATO

La teoría y práctica de los metadatos tiene su origen en la catalogación de documentos impresos y consiste en incorporar texto para describir los atributos de los documentos, lo cual los dota de un significado, un contexto y una organización. La creación e implementación de los metadatos es un proceso intensivo que requiere de tiempo, recursos humanos y económicos, por lo que es necesario hacer un análisis de los costos y beneficios de su uso considerando las necesidades de los usuarios y los gestores de la información.

Los metadatos son por tanto una herramienta administrativa que da datos sobre los datos, y proporciona información sobre objetos digitales y no-digitales que se pueden procesar por una computadora y ser interpretados por el ser humano.

Categorías de Metadatos

- ❖ Descriptivos
- ❖ Administrativos
- ❖ De preservación
- ❖ Estructurales
- ❖ Meta-metadatos
- ❖ Otros

METADATOS DE PRESERVACIÓN

Los metadatos, como elementos para la descripción de los documentos digitales o DLO⁶ deben aportar información sobre las características lógicas a que se refieren para su correcta preservación, así como contener información sobre el desarrollo adecuado de actuaciones

6 Document like object.

III Encuentro de Catalogación y Metadatos

de preservación al aportar información sobre el tipo de formato, modificaciones realizadas sobre éste y otras informaciones relevantes.

Cabe señalar que el proceso de gestión de la preservación digital necesita ser automatizado y basarse en metadatos que sean fácilmente interpretados. Por lo tanto al ser los metadatos información digital, su preservación debe ser tomada en cuenta.

MODELOS DE METADATOS EN DESARROLLO

Dublin Core Metadata Initiative Preservation Working Group

Establecido en 2003, Dublin Core funciona como un grupo de discusión y debate para personas y organizaciones interesadas o que trabajen sobre metadatos para la preservación. Sus objetivos son: recopilar información y hacer un estudio de los esquemas de metadatos existentes, así como investigar la necesidad de tener un campo específico para el esquema de metadatos de preservación y establecer contacto con otros proyectos globales que tengan el mismo propósito.

Los datos recogidos en el modelo Dublin Core se agrupan en tres tipos: aspectos relacionados con el contenido del DLO; con la propiedad intelectual del recurso, y los vinculados a la temporalidad del documento: fecha, tipo, formato e identificador del recurso. Este último grupo debe contener las instrucciones necesarias para el tratamiento correcto del DLO (programa/s utilizado/s para su creación, lectura, modificación, entre otros y los que contengan el historial de los posibles cambios sufridos de formato, soporte, etcétera).

Premis Working Group Patrocinado por OCLC y RLG (Research Libraries Group)

Sus objetivos son desarrollar el núcleo para el conjunto de metadatos de conservación, apoyándose en un diccionario de datos que tienen una amplia aplicabilidad por la comunidad interesada en este

tema, e identificar y evaluar estrategias para codificar, almacenar y gestionar metadatos en sistemas de preservación digital.

Desde mayo de 2005 están disponibles las conclusiones y resultados del trabajo en el *Data dictionary for preservation metadata: final report of the Premis working group*. El documento define y describe un núcleo de metadatos de preservación con amplia aplicabilidad en repositorios, y además analiza un esquema de metadatos y sus diferentes elementos, así como indicaciones sobre su uso y da ejemplos de aplicación a distintos tipos de DLO (documentos de MS Word, páginas web, fotografías, etcétera). Intenta presentarse como una base sobre la cual desarrollar futuros esquemas.

El modelo OAIS⁷

Un estándar de ISO⁸ propone un marco de referencia para preservar colecciones digitales, se trata de un sistema abierto que presenta un nuevo enfoque sobre la función de preservación de los metadatos.

Destaca el ISO cinco categorías de metadatos:

1. Referencia: define los identificadores o códigos de referencia asignados a la información e identifica de forma unívoca al archivo, tanto a nivel interno como externo.
2. Contexto: documenta la relación de la información con el entorno.
3. Procedencia: documenta la historia de la información: sus orígenes, historia de la custodia, intervenciones para garantizar la preservación, integridad y autenticidad de la información.
4. Autenticación de la información: son los mecanismos de autenticación que garantizan que la información contenida no se ha alterado (comprobación, firma digital, etcétera).
5. Representación de la Información: se encarga de revisar los enfoques existentes en metadatos de preservación.

7 Open Archival Information System.

8 ISO 14721:2003.

III Encuentro de Catalogación y Metadatos

El modelo OAIS puede considerarse como un punto de partida. El modelo establece la terminología y los conceptos relevantes para los archivos digitales, e identifica los principales componentes y procesos, a la vez que propone un esquema de información para los objetos digitales y los metadatos asociados. No hace referencia a la implementación ni a estrategias específicas de preservación y por tanto mantiene una posición neutral en cuanto a una tipología de objetos digitales o aspectos tecnológicos, lo cual es aplicable a una amplia variedad de instituciones que trabajan con objetos digitales.

A partir de este marco se han desarrollado otras iniciativas como es el caso de los proyectos que describimos a continuación:

CURL Exemplars in Digital Archives (CEDARS)

Éste fue concebido a partir de dos objetivos: el primero, ser un esquema para el archivo digital piloto de CEDARS, y el segundo, ser una contribución para desarrollar un sistema estándar de metadatos de preservación. En este sistema los metadatos permiten la preservación a largo plazo de recursos digitales y resultan necesarios para acceder al contenido digital incluyendo información descriptiva, administrativa, técnica, y legal. Los metadatos propuestos pueden ser aplicables a una amplia gama de objetos digitales, pues el sistema considera a todos los formatos de documentos que puedan estar presentes en una biblioteca. Puede ser aplicable con independencia del nivel de especificidad que se le asigne el metadato

Nacional Library of Australia (NLA)

El enfoque utilizado en los metadatos propuestos, está dirigido fundamentalmente a la información para gestionar la preservación. Esta propuesta esta motivada por la necesidad de preservar tanto objetos digitales originales como “sustitutos” o copias digitales. Aborda además, la necesidad de contar con diferentes niveles de descripción: colección, objeto y sub-objeto, aunque reconoce el objeto como el principal elemento de la descripción.

Networked European Deposit Library (NEDLIB)

Maneja un número mínimo de metadatos que son obligatorios para gestionar la preservación y permitir el manejo de grandes cantidades de datos en un contexto tecnológico cambiante. El sistema está pensado para abordar la obsolescencia tecnológica, aspecto que es considerado el principal objetivo de los metadatos de preservación. Intenta lograr esto centrándose en la información de carácter genérico de los objetos digitales, y su objeto es favorecer su aplicación para poder ser empleado por una amplia gama de formatos y tipos de objetos.

CONCLUSIONES

- ❖ No basta con digitalizar información es necesario establecer los mecanismos para establecer políticas de preservación digital que hayan sido definidas por reglas y procedimientos. De no hacerlo así, gran parte de la información se perderá en la siguiente década.
- ❖ Es contradictorio que los procesos para mejorar el desempeño de las TIC, como mejorar la capacidad de compresión y avanzar en el almacenamiento masivo, genere problemas de incompatibilidad al acelerar la obsolescencia tecnológica.
- ❖ Los procesos de preservación deben dar respuesta a problemas como el rejuvenecimiento del material, la migración y la preservación.
- ❖ Será determinante al planear proyectos de digitalización, trabajar bajo directrices que converjan con otros proyectos desarrollados con el fin de compartir entre ambas información; por ello el cuidado que una organización invierta al comienzo del proceso en el uso de estándares, documentación, gestiones de los ficheros y otras prácticas, se verá realizado en menores requisitos de preservación y mantenimiento.
- ❖ El desarrollo de un marco general de metadatos de preservación contribuirá con la preservación futura de los archivos digitales además de compartir recursos y de facilitar el intercambio de metadatos. Sin embargo la diversidad de materiales digitales, de

instituciones implicadas y de contextos de aplicación existentes ha dado pie a numerosas propuestas y esquemas de metadatos. Esta situación dificulta la creación de un esquema único para la descripción de los metadatos, sin embargo sería deseable el desarrollo de un marco común que satisfaga diferentes puntos de vista y provea un conjunto de métodos y herramientas que faciliten la incorporación de los generadores de contenido en la gestión de la preservación digital y en el proceso de creación de los metadatos.

BIBLIOGRAFÍA

- A METADATA Framework to Support the Preservation of Digital Objects, [en línea], http://www.oclc.org/research/projects/pmwg/pm_framework.pdf (25 mar 2008).
- Candas Romero, Jorge, “El papel de los metadatos en la preservación digital”, en *El Profesional de la Información*, v.15, núm. 2, marzo-abril, 2006, p. 126-136.
- Day, Michael, Metadata for digital preservation: an update, [en línea], <http://www.ariadne.ac.uk/issue22/metadata> (8 may2008)
- Deegan, M., Tañer, S.: 8, “Preservation. Digital Futures series”, en *Digital Futures: strategies for the information age*, Library Association Publishing, 7 Ridgmount Street, London, WC1E 7AE (2002), 179-208.
- Guinchat Claire, Introducción general a las ciencias y técnicas de la información y documentación, España, CINDOC, 1990.
- Harvey, Ross, *Preservation in libraries: principles, strategies and practices for librarians*, London, Bowker Saur, 1993, (Topics in Library and Information Studies).
- Martínez Comeche, Juan Antonio, Teoría de la información Documental y de las instituciones documentales, España, Síntesis, 2000.

Los metadatos, una herramienta para preservar al conocimiento digital

Mayor Zaragoza, Federico, “Sociedad de la Información en el siglo XXI: un requisito para el desarrollo II Reflexiones y conocimiento compartido, Capítulo 7”, [en línea], http://www.desarrollosi.org/Volumen2/Web/PDF/Cap7_La_persona_en_la_SI.pdf (10 ene 2007).

Moreira González, José Antonio, “Nuevas Competencias profesionales para nuevas funciones”, [en línea], <http://www.anabad.org/archivo/docdow.php?id=204> (14 dic 2006).

Peig Olivé, Enric, Interoperabilidad de metadatos en sistemas distribuidos, Barcelona, Universitat Pompeu Fabra, 2003, [en línea]. http://www.tdx.cbuc.es/TESIS_UPF/AVAILABLE/TDX-0316104-132946/tepo1de1.pdf (10 abr 2008).

Ros García Juan, Políticas de Información y documentación, España, Síntesis, 2002.

Sánchez Manuel, Desarrollo de una política de preservación digital, España, Grijalbo, 2004.