



---

# Algunos apuntes sobre la historia de los metadatos de preservación para objetos digitales en el contexto de las bibliotecas

Some notes on the history of preservation metadata for digital objects in the context of libraries

Luis Omar Méndez Saavedra

Estudiante Graduado  
Escuela Graduada de Ciencias y Tecnologías de la Información  
Universidad de Puerto Rico, Recinto de Río Piedras  
[luis.mendez1@upr.edu](mailto:luis.mendez1@upr.edu)

Recibido: febrero 2020 | Aceptado: mayo 2020

---

## Resumen

Este trabajo consiste en una revisión de la literatura publicada sobre los metadatos, con un enfoque en el tipo de metadatos utilizado para la preservación de objetos en bibliotecas y repositorios digitales. Se discute la evolución a través de los años de los metadatos de preservación, la importancia de estos para las labores de bibliotecarios, manejadores de datos y archivistas, al igual que su importancia para asegurar la durabilidad a largo plazo de los documentos digitales. Se presentan distintas perspectivas en torno a la función fundamental de este tipo de metadatos, el rol que estos juegan y áreas de énfasis específicas a disciplinas. Además, se discuten algunos de los esquemas de metadatos que han sido creados específicamente para dar soporte al registro de metadatos de preservación de más amplio uso en la actualidad.

*Palabras clave:* metadatos, preservación, preservación digital, estándares de metadatos, biblioteca digital

## Abstract

This article presents a literature review on the subject of metadata, specifically on the type of metadata used for the preservation of objects in libraries and digital repositories. It discusses the evolution of preservation metadata, its importance for librarians,

archivists and data stewards, and its key role in ensuring the long-term preservation of digital documents. The article compares different perspectives on the function and role of preservation metadata and discusses various metadata schemes.

*Keywords:* metadata, preservation, digital preservation, metadata standards, digital library

## **Algunos apuntes sobre la historia de los metadatos de preservación para objetos digitales en el contexto de las bibliotecas**

El enfoque principal de este trabajo son los metadatos para la preservación digital debido a su importancia y particular relevancia en nuestro contexto informático contemporáneo. El continuo y acelerado desarrollo tecnológico de nuestra época ha exacerbado la necesidad para la creación de mecanismos capaces de preservar la inmensa cantidad de información producida en nuestros quehaceres laborales y cotidianos, de manera que esta sea recuperable y accesible en el futuro. En la actualidad, la creación de objetos nativos al ambiente digital se ha convertido en la norma más bien que en la excepción, debido al ya mencionado desarrollo tecnológico y la fluidez en la transmisión de información que posibilita su naturaleza digital. El objetivo primordial de este trabajo es elaborar un panorama histórico en torno a la discusión que se ha establecido sobre la necesidad de preservar recursos digitales, el rol protagónico de los metadatos para este propósito, las diversas maneras que han sido propuestas para enfrentar este reto y el estado actual de la cuestión. La metodología empleada para este trabajo consistió en la recopilación de diversas fuentes informativas relevantes al tema, su discusión y análisis con el fin de presentar una visión retrospectiva y panorámica del asunto y proponer alguna ruta posible a seguir.

### **Revisión de la literatura y desarrollo**

La preocupación en torno al manejo responsable y la preservación adecuada de los objetos de información o DLO (*document like objects*) digitales ha estado presente en los círculos de los profesionales de la información desde los albores de la era digital. El surgimiento de un nuevo soporte para la información supuso un reto disciplinar sin precedentes para estos profesionales, puesto que la llegada de la era digital constituyó una revolución en la manera en la cual se difunde la información y cómo interactuamos con los recursos informativos que sobrepasó incluso la invención de la imprenta europea. La naturaleza siempre cambiante de las tecnologías a partir de las cuales son creados los objetos digitales constituye uno de los mayores desafíos para lograr su preservación pues, como señalan Corrado & Moulaison (2017), “As information continues to go digital and new technologies are developed, formats that are older or outdated become obsolete and content may not be accessible using new software” (p.8). Al enfrentarse a este nuevo reto, muchos de los profesionales de la información optaron por idear maneras para



adaptar los conocimientos disciplinares tradicionales a esta nueva realidad carente de materialidad.

Candás (2006) cita el planteamiento de Jeff Rothenberg en torno a cómo abordar la cuestión de la preservación de documentos digitales partiendo de una definición para la preservación digital arraigada en la preservación tradicional de documentos:

La preservación de documentos digitales es análoga a la de textos antiguos. Al igual que con los documentos digitales, a veces es necesario actualizar un texto antiguo transcribiéndolo[...]. Un texto antiguo puede ser preservado de dos modos: bien copiándolo en su idioma original o bien traduciendo a cualquier otro idioma contemporáneo al momento de la transcripción (p.128).

La preservación digital es una labor que conlleva un

conjunto de actuaciones relativas a la preservación de los documentos existentes en formato digital. [...] destinadas a que un objeto perdure el mayor tiempo posible en su estado original [...] preocupándonos no solo por el mantenimiento del objeto, sino también (y principalmente) por su contenido informativo (Candás, 2006, p.128).

La obsolescencia tecnológica es uno de los mayores peligros a los que se enfrentan los objetos digitales. A propósito de combatir la misma se han propuesto al menos tres alternativas para preservar la usabilidad de los objetos digitales: la migración del soporte físico y/o de los formatos de archivos a nuevas versiones de estos, la emulación del soporte antiguo mediante *software* utilizando un equipo nuevo y la preservación de la tecnología (*hardware* y *software*) necesaria para mantener legible la información en su soporte original (Candás, 2006, p.129).

### **Sobre metadatos, OAIS y PREMIS**

Para realizar el tipo de tarea requerida para la preservación digital, es necesario contar con la información sobre el formato original del objeto, en este ejemplo, para lograr el proceso de conversión. En la labor de servir este propósito, y muchos otros, es menester la utilización de metadatos. Los metadatos son comúnmente definidos como “datos sobre datos”. Lazinger (2001) cita otro texto que provee una definición preliminar más abarcadora, la cual define algunas de las funciones que cumplen los metadatos:

The term “metadata” commonly refers to any data that aids in the identification, description and location of networked electronic resources. A primary function of metadata is resource discovery: metadata increases the odds that a user will be able to retrieve appropriate information and assess its usefulness and reliability.

Another important function provided by metadata is control of the electronic resource, whether through ownership and provenance metadata for validating



information and tracking use; rights and permissions metadata for controlling access; or content ratings metadata, a key component of some Web filtering applications (p.140).

Ha habido diversos acercamientos hacia cómo dividir y clasificar los distintos tipos de metadatos utilizados en la bibliotecología digital. Algunos autores identifican tres tipos generales, otros hasta cinco. Corrado y Moulaison (2017) identifican cuatro tipos principales de metadatos: metadatos descriptivos, metadatos administrativos, metadatos técnicos y metadatos estructurales (p.65). Cada tipo de metadato recoge información particular sobre los objetos digitales y, en conjunto, aumentan las posibilidades para que éstos sean descubiertos, se haga más fácil el acceso y aumente su perdurabilidad. El tipo de metadatos que concierne este escrito son los metadatos para la preservación digital. Es imposible precisar cuáles metadatos específicos forman parte de este tipo, puesto que distintos metadatos pertenecientes a cualquiera de los cuatro tipos de metadatos mencionados anteriormente pueden ser, en efecto, metadatos de preservación, ya que la información que se debe registrar y utilizar para asegurar la preservación es amplia y abarcadora.

El *Open Archival Information System* (OAIS) es un modelo de referencia pionero en visibilizar la necesidad de establecer métodos de preservación para los objetos digitales y proponer una visión general sobre el tipo de información, registrada en forma de metadatos, requerida para este propósito. OAIS identifica un *Content Information* (CI) como “a set of information that is the original target of preservation or that includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information” (Corrado y Moulaison, 2017, p.177). Esto significa que el CI se refiere al conjunto del objeto digital y sus metadatos que deben ser preservados. El modelo OAIS hace hincapié en la necesidad de proveer información que permita asegurar la autenticidad del *CI object* como parte de los metadatos para la preservación digital. En función de esto, OAIS crea y define el concepto de *Preservation Description Information* (PDI) como “the information [that] is necessary for the adequate preservation of the Content Information and [that] can be categorized as Provenance, Reference, Fixity, Context and Rights Information” (Corrado y Moulaison, 2017, p.178). Brevemente, la información de procedencia es la que documenta la historia del CI, la información de referencia se utiliza como identificador único del CI, la información de estabilidad (*fixity*) documenta los mecanismos utilizados para asegurar que el *CI object* no haya sufrido alteraciones indocumentadas, la información de contexto documenta las relaciones del CI con su entorno y la información sobre derechos de acceso identifica las restricciones impuestas al acceso del CI (Corrado y Moulaison, 2017, p.180).

Estos autores afirman que los metadatos para la preservación digital deben incluir información que dividen en tres amplias categorías: los metadatos necesarios para el almacenamiento y recuperación en el ámbito digital (descriptivos, administrativos, técnicos, y estructurales), la PDI para asegurar la autenticidad y metadatos sobre los objetos digitales y su entorno digital actual, lo cual permite su preservación y acceso a



largo plazo a través del sistema de preservación digital (Corrado y Moulaison, 2017, p.182). El modelo PREMIS (*Preservation Metadata: Implementation Strategies*) es actualmente el principal esquema para metadatos de preservación. El grupo de trabajo que creó este modelo también publicó en el 2005 la primera versión del *PREMIS Data Dictionary*, hoy día en su tercera versión actualizada. Un diccionario de metadatos como este describe:

1. Significados comunes de metadatos (semántica).
2. Gramática y reglas comunes para expresar los datos (sintaxis).
3. Propiedades de elementos de diccionarios de metadatos comúnmente definidas (atributos).

PREMIS fue creado originalmente con la intención de ser utilizado con el modelo OAIS como referencia. Sin embargo, a través de sus actualizaciones, se ha convertido en algo más cercano a un estándar de manejo del ciclo de vida ya que ha incorporado información utilizada para sustentar usos más tradicionales de metadatos, además de los de preservación (Corrado y Moulaison, 2017, p.184). PREMIS define cuatro entidades: entidad objeto, entidad evento, entidad agente y entidad declaración de derechos. La entidad objeto se divide en cuatro categorías: entidad intelectual, entidad representación, entidad expediente y entidad *bitstream*. A estas se les atribuyen unidades semánticas, definidas como “propiedades de una entidad” (p.184). Aunque el rol de estas parece similar al de los elementos de metadatos en otros esquemas, el diccionario de datos establece una clara diferencia sobre su naturaleza. Las unidades semánticas son la información que un repositorio de preservación necesita saber, mientras que los elementos de metadatos son la manera en que esa información es grabada.

El modelo PREMIS es altamente flexible y acomodable, de manera que permite ser complementado y/o complementar otros esquemas de metadatos al ser utilizados en conjunto. El *Metadata Encoding and Transmission Standard* (METS) es un estándar de metadatos creado con el propósito de guardar los metadatos estructurales del recurso - la relación que tienen partes del recurso entre sí o con otros recursos- que puede incorporar el sistema de unidades semánticas provistas por el *PREMIS Data Dictionary* para almacenar los metadatos de preservación necesarios para los recursos. También existen esquemas de metadatos creados para registrar tipos de metadatos más específicos. El *DigiProv<sup>1</sup> Extension Schema* fue creado por la Biblioteca del Congreso a propósito de ser utilizado para registrar los metadatos de procedencia. Este cuenta con un diccionario de datos que identifica elementos y subelementos para la configuración de expedientes, instancias, procesos, ajustes, tareas y herramientas. El *Technical Metadata for Text* (textMD) es un esquema mantenido por la Biblioteca del Congreso

---

<sup>1</sup> Digital Provenance



utilizado para registrar propiedades como la información de codificación, información del tipo de caracteres e idiomas, entre otras (Corrado y Moulaison, 2017, p.75-77).

## Otros estándares

Candás (2006) discute algunos proyectos que se estaban realizando a la fecha de publicación de su texto, entre ellos el ya discutido PREMIS. Otro de los proyectos, *The Cedars Project*, desarrolló una propuesta para un esquema de metadatos cuyo eje principal era asegurar “el acceso significativo a objetos digitales durante décadas o incluso siglos precisara de un sofisticado sistema de metadatos que describan detalladamente procesos técnicos, de gestión y de actividades administrativas, así como de control de acceso y copyright”. Con este fin, la propuesta del *Cedars Project* define tres niveles de metadatos clasificados de acuerdo con el nivel de importancia que tienen los elementos de descripción (Candás, 2006, p.133):

- Muy significativos: entre ellos se encuentran: *preservation description information, reference information, context information, provenance information, custody history, management history, rights management.*
- Significativos: *entre ellos se encuentran: existing metadata, reason for creation, original technical environments, prerequisites, reason for preservation.*
- Poco significativos: se consideran útiles, pero no estrictamente necesarios. No se provee ejemplo de los metadatos a este nivel.

La evidente importancia que esta propuesta le otorga a los metadatos de preservación en función de precisar la autenticidad se debe a que estaba basada, en gran medida, en el modelo OAIS, incluso en la estructura y los apartados que propone para su esquema:

- Información de referencia
- Información de contexto
- Información de procedencia
- Información de estabilidad (*fixity*)
- Información de representación
- Derechos de propiedad intelectual y uso

El otro proyecto que se menciona fue uno de un esquema propio creado por la Biblioteca del Congreso para la descripción de repositorios digitales agrupando sus metadatos en tres tipos: descriptivos, administrativos y estructurales. El vocabulario para este esquema estaba basado en el *Structural metadata dictionary for LC digitized material, version 1.03*. Los metadatos definidos por este esquema deben incluir información para la preservación digital e información para el cambio de formatos, denominados como *preservation (D)* y *preservation (R)*, respectivamente. Entre los elementos para los primeros están: *archive\_date\_time, archive\_history, archive\_ID,*



*archive\_next\_data\_time* y *associated\_file\_type*. Los segundos incluyen: *reformatted\_original\_information* y *reformatting\_guidelines* (Candás, 2006, p.134).

Hasta este punto la discusión en torno a la importancia de los metadatos para la preservación de objetos digitales ha sido presentada, principalmente, desde el punto de vista de la bibliotecología y el manejo de la información. Otras disciplinas, como la archivística moderna, también tienen mucho que aportar en cuanto al tema que nos concierne en este texto. Wilson (2010) presenta el concepto del *records continuum model*, desarrollado en Australia durante los 1980s y 1990s, aplicado al manejo de registros en la era poscustodial. El valor de este concepto para nuestro interés radica en que nos permite pensar sobre los objetos digitales de una manera más holística, reconociendo la fluidez que caracteriza la existencia de estos objetos y socavando los límites artificiales inherentes a los modelos del ciclo de vida tradicionales (Wilson, 2010, pp. 207-208). Es desde esta conceptualización de los registros/objetos digitales que se reconoce la importancia de los metadatos para la retención a largo plazo de estos en la archivística digital. En reconocimiento de esta necesidad se formulan los metadatos para el mantenimiento de registros (*recordkeeping*), cuyo propósito primordial es asegurar que un registro sea utilizable a lo largo del tiempo por medio de proveer información estandarizada sobre la identidad, autenticidad, contenido, estructura, contexto y requisitos esenciales para el manejo de registros. El propósito fundamental de estos metadatos es asegurar la confiabilidad, autenticidad, usabilidad e integridad de los registros digitales (Wilson, 2010, pp.208-209). Este tipo de metadatos es, prácticamente, la versión de metadatos para la preservación digital propia a la disciplina archivística.

La concepción de una nueva clase de metadatos conlleva la necesidad de crear formas estandarizadas para su correcta implementación. En el 1999 fueron creados dos estándares de metadatos para el mantenimiento de registros (*recordkeeping metadata*). El primero de estos fue publicado por la National Archives of Australia (NAA) y definía distintos roles para las propiedades de los metadatos: de registro, de términos y condiciones, estructurales, contextuales, de contenido e historia de uso. Por otra parte, la Universidad de Monash desarrolló su propio estándar, con grandes similitudes a estándares más recientes como el de PREMIS. El modelo Monash definió cuatro clases principales de entidades (*business entities, agent entities, records entities, business-recordkeeping entities*) y le adjudicó una serie de elementos de metadatos para cada entidad, incluyendo diez elementos en común y otros específicos a la descripción de cada entidad. El estándar más reciente es el *Australian Government Recordkeeping Metadata Standard* (AGRkMS), actualmente vigente y en su versión 2.2. Este modelo es una síntesis de los dos estándares previamente mencionados, añadiendo complejidad al primero y expandiendo el número de entidades del segundo a cinco (*agents, records, business, mandates, relationships*) de manera que sea capaz de describir detalladamente las propiedades de los metadatos sobre registros y contextos de creación para asegurar una rigurosa preservación de estos. (National Archives of Australia, 2015)



Dappert y Enders (2010) destacan LMER (*Long-term preservation Metadata for Electronic Resources*) de la Biblioteca Nacional de Alemania como ejemplo para una alternativa a la homogeneidad del modelo PREMIS. En contraste directo con el modelo PREMIS, LMER fue diseñado a propósito de usarse para un proyecto específico. El mismo incorpora distintas estrategias de preservación, la habilidad de documentar la procedencia de objetos digitales, metadatos técnicos básicos, entre otras cosas (Dappert y Enders, 2010, p.9). La importancia que tiene un modelo local como LMER radica en que permite la creación de modelos que están mejor adecuados a las necesidades particulares y de acuerdo con la naturaleza del trabajo realizado en la institución. Sin embargo, se debe mantener al mayor grado posible la conformidad con estándares y reglamentos internacionales del modelo diseñado con necesidades particulares en mente de manera que sea posible la interoperabilidad de este con otros.

### **Conclusiones (inconclusas)**

A lo largo de este trabajo realizamos un resumen panorámico sobre lo que son los metadatos y su vital importancia en nuestra era de información digital. Examinamos algunas de las maneras y razones por las que los metadatos para la preservación digital fueron cobrando popularidad entre profesionales de la información preocupados por la cuestión de mantener la accesibilidad, la integridad y permanencia a largo plazo para sus objetos digitales. Luego, discutimos algunos modelos y esquemas que han sido formulados en función de definir lo que son los metadatos para la preservación digital, cómo clasificarlos, cómo organizarlos y distintas maneras para registrarlos. Entre los esquemas que tratamos vimos algunos que eventualmente terminaron cayendo en desuso, o bien que nunca pudieron despegar, ante el auge de PREMIS y su imposición como el estándar de facto para la incorporación de metadatos para la preservación de objetos digitales. Sin embargo, la flexible adaptabilidad que permite PREMIS para funcionar con otros esquemas constituye un punto a su favor en cuanto a sus posibilidades futuras. Además, como pudimos ver, aún quedan alternativas que retan al modelo PREMIS y lo empujen hacia adelante de manera que no se estanque y continúe progresando y actualizándose para servir las nuevas necesidades de preservación digital que inevitablemente surgirán en el futuro debido al continuo desarrollo de las tecnologías digitales. Sin duda alguna los metadatos de preservación continuarán siendo de vital importancia, no solo para los profesionales de la información encargados de emplearlos para mantener los recursos digitales, sino también para la comunidad investigadora y el público en general que se beneficia del resultado de la labor de preservación.



## Referencias

- Candás Romero, Jorge (2006). El papel de los metadatos en la preservación digital. *El profesional de la información*, 15(2), 126-136.
- Corrado, E. M. & Moulaison, H. (2017). *Digital preservation for libraries, archives and museums*. Rowman & Littlefield.
- Dappert, A. & Enders, M. (2010). Digital preservation: metadata standards. *Information Standards Quarterly*, 22(2), 5-13. doi: 10.3789/isqv22n2.2010.01
- Groenewald, R. & Breytenbach, A. (2011). The use of metadata and preservation methods for continuous access to digital data. *The Electronic Library*, 29(2), 236-248. doi: 10.1108/026404711111125195
- Lazinger, S. S. & Tibbo, H. R. (2001). *Digital preservation and metadata: History, theory and practice*. Libraries Unlimited.
- Liu, Jia. (2007). *Metadata and its applications in the digital library: Approaches and practices*. Libraries Unlimited.
- National Archives of Australia & Archives New Zealand. (2015). Australian Government Recordkeeping Metadata Standard (AGRkMS). [Canberra]: National Archives of Australia. Available at: <https://nla.gov.au/nla.obj-370643959/view>
- Wilson, A. (2010). How much is enough: metadata for preserving digital data. *Journal of Library Metadata*, 10, 205-217. doi: 10.1080/19386389.2010.506395
- Zeng, M. L. (2008). *Metadata*. Neal Schuman Publishers.

