

# Metadatos para preservación digital

JUAN VOUTSSÁS MÁRQUEZ

Universidad Nacional Autónoma de México

**L**os metadatos existen desde mucho antes de que se acuñara este nombre en la década de los sesenta. Se encontraban ya en forma de descriptores de contenido en tabletas de arcilla de la biblioteca de Ebla en el tercer milenio antes de nuestra era, en la Biblioteca de Alejandría, a lo largo de la Edad Media y, por supuesto, en la edad contemporánea. Debido a esa antigüedad y a que siguen siendo sin duda de la mayor importancia, han sido los más tratados y desarrollados y se les presta mucha atención. Hoy en día existe un sinnúmero de definiciones, métodos, variantes, teorías, estándares, etcétera, relativos a los descriptores de contenidos. No obstante, existen muchos otros atributos de los documentos que también pueden ser representados con metadatos.

Este texto pretende tratar en especial aquellos metadatos para preservación de documentos digitales. Su importancia radica en el hecho de que con mucha frecuencia pasan desapercibidos y son ignorados por los responsables de esos documentos, lo cual pone en riesgo su adecuada preservación. Sin adecuados metadatos creados expresamente al efecto, será muy difícil, o casi imposible, lograr la preservación ulterior de un documento digital.

La preservación de documentos digitales dentro de un fondo documental, llámese esta colección, repositorio, acervo, archivo, etcétera, es un proceso que debe ser diseñado al momento de concebir ese fondo. Muchos todavía piensan que la preservación de documentos digitales es exclusivamente un problema tecnológico, y que por eso su solución consiste en la adquisición de dispositivos y programas computacionales y algunos procesos y accesorios para respaldos de la información, y por tanto, es responsabilidad total del área de informática de la organización.

En consecuencia, el momento en que se comienza a pensar en la preservación es cuando el fondo documental está creado, estable y lleva buen tiempo operando. Ésa es una receta para el desastre. Las políticas, los procedimientos y las estructuras para la preservación documental digital se deben diseñar y construir al mismísimo inicio del fondo documental para que cuando este comience a construirse, ya contenga todos los elementos que permitirán a los documentos digitales ser preservados por todo el tiempo necesario. Pretender agregar después estos componentes de preservación es prácticamente imposible y cuando llega a hacerse, implica una enorme inversión de recursos económicos, técnicos y humanos. Es por tanto indispensable diseñar la preservación de inicio junto con el fondo documental correspondiente.

Para comenzar, conviene establecer que se preservan objetos digitales y no simples documentos digitales. De forma simple, puede definirse que

[...] un documento digital es una unidad de información persistente constituida por un mensaje fijado a un medio de manera sintácticamente estable, cuyos datos están representados por medio una sucesión de valores numéricos diferenciados, por lo

general secuencias de bits, de acuerdo con un código o convención preestablecido (Voutssás 2017, 206).

Esta definición puede ser complementada agregando que esa unidad de información puede contener cualquier tipo de ella: texto, imágenes, sonido, video, y sus combinaciones, en cualquiera de sus formas de expresión conocidas: libros, revistas, artículos, tesis, diarios, documentos de archivo, mapas, partituras, gráficas, música, cine, juegos, programas y aplicaciones informáticas, correos electrónicos, arte digital, etcétera. No obstante, conviene que este concepto de “documento digital” sea ampliado al de “objeto digital”. Existen muchas definiciones al respecto; por ejemplo, la Biblioteca Digital de California establece en su glosario: “Objeto digital: entidad en la que uno o más archivos de contenido y sus metadatos correspondientes están unidos físicamente y/o lógicamente mediante el uso de un contenedor digital”.

Habiendo algunas sutiles diferencias entre ambos conceptos relacionadas sobre todo al contexto, un objeto digital, para resumirlo de forma simple, es:

[...] un objeto documental representado numéricamente por una agregación o secuencia de valores numéricos en forma de una o más cadenas de bits las cuales –bajo ciertas reglas de representación– contienen datos dentro de ese objeto, así como metadatos acerca de las propiedades del mismo y –cuando sea necesario–, los métodos para realizar operaciones sobre el objeto. El significado de las cadenas de bits como objeto documental está en función de la interpretación que hagan de ellas los programas de su entorno digital (Interpares s.f.).

Puede verse que las definiciones de “documento digital” y “objeto digital” son semejantes; la diferencia fundamental

consiste en que en las definiciones de objeto digital se aclara que estos tienen, además de los datos o contenido, metadatos y eventualmente procedimientos. Básicamente, quiere decir que un documento digital es un contenido, mientras que el objeto digital incluye otros elementos más allá del contenido, y en esos elementos no pueden faltar los metadatos.

Conviene también establecer que para fines de este texto, la *preservación digital* tiene que ver con el contenido del documento, y su alcance siempre es el largo plazo; trata objetos lógicos. Como complemento, la *conservación digital* tiene que ver con el soporte y forma del documento digital, y su alcance es siempre el corto-mediano plazo; trata objetos físicos, es sinónimo de *mantenimiento*. Como corolario, podemos afirmar que conservamos objetos físicos (soportes) para preservar objetos lógicos (contenidos). *Preservación digital* es a la vez un objetivo y un resultado: como objetivo, “preservación digital” es en esencia “asegurar a largo plazo de la permanencia y acceso del contenido de documentos digitales confiables a lo largo del tiempo y las tecnologías”. Al instrumentar y lograr este objetivo, se obtiene como resultado –se logra– la “preservación digital” de los materiales; esto es: se obtiene “la protección del contenido intelectual de materiales digitales –datos, documentos o archivos” (Voutssás y Barnard 2014).

Para que un objeto digital se preserve, debe mantener a lo largo del tiempo ciertas características inherentes. Éstas varían de acuerdo a cada tipo de documento, así como su importancia y peso específico, pero en general pueden establecerse seis características principales (Voutssás 2009):

- Permanencia (la información existe). Esta característica está relacionada con el concepto de que el documento estará disponible por un lapso

considerable. Está asociado con su presencia, su seguridad y, por supuesto, con la duración y continuidad de su soporte. La permanencia depende en mucho del almacenamiento permanente seguro de los objetos digitales. Para ello son necesarios soportes, procedimientos y técnicas adecuadas para mantener esos objetos a largo plazo conservando las *cadena de bits* que forman el documento digital para asegurar la presencia de esos datos. Si el soporte o los datos no permanecen, el documento ya no estará disponible.

- **Accesibilidad** (la información es reinterpretable y usable). Tiene que ver con la capacidad de poder acceder en un futuro a la información o datos contenidos en un cierto soporte. El hecho de lograr que el soporte y las cadenas de *bits* contenidas en él se conserven en buen estado no implica que podrán ser utilizables. Se debe poder asegurar que habiéndose conservado en buen estado se podrán ver o ejecutar estos archivos informáticos dentro de mucho tiempo reproduciendo correctamente las cadenas de *bits*. Si el soporte o formato del fichero informático pertenecen a tecnologías muy obsoletas y ya no pueden ser reproducidos porque se carece del equipo para leerlos o del sistema operativo o el programa informático que los operaba, la información existe, pero no puede ser accedida a ella. Independientemente del buen estado de los datos, es decir, de que haya habido *permanencia*, puede o no haber *accesibilidad*. Como puede verse, son dos características distintas

pero complementarias; se requiere que existan ambas para un adecuado acceso. La falta de cualquiera de los dos hará inaccesible al documento en cuestión.

- Disponibilidad (la información puede ser encontrada y accedida). Tiene que ver con que el documento está alojado en un soporte que seguirá siendo accesible a través del tiempo, y que la información está debidamente registrada en algún sistema de búsqueda para asegurar que podrá ser localizada y reproducida.
- Confidencialidad y privacidad. Se refiere al hecho de que el documento solo pueda ser visto o utilizado por las personas autorizadas para ello, y que los datos personales asociados al documento no puedan ser transferidos a terceros no autorizados ni extraídos por ellos.
- Autenticidad (atributos para que el documento ser considerado original y fiable). Cada tipo de documento debe tener previamente establecidos aquellos atributos que a lo largo del tiempo siga siendo considerado auténticos y veraces (autoría, identidad, integridad, estabilidad, contexto, custodia, forma fija, etcétera) y que, por tanto, pueda seguir siendo considerado confiable, aun a pesar de cambios que se hayan introducido al documento esporádicamente debido a actualizaciones tecnológicas.
- Aceptabilidad (no repudio). Tiene que ver con las características que permiten a un cierto documento seguir siendo aceptado por el emisor y los usuarios: números DOI, ISBN; firmas, sellos, y marcas de agua electrónicas; *hashings*, *checksums*,

depósito en repositorios oficiales fiables, cifrados, etcétera.

Si se crean de inicio y se mantienen, todas estas características en conjunto mantendrán la usabilidad y confiabilidad de la información de un cierto documento a lo largo del tiempo. Para saber que estos atributos existen y que se han conservado a lo largo del tiempo, se requiere indispensablemente de dotar al documento de metadatos específicos para ello que, como puede intuirse, no tienen ya que ver con la descripción de contenido del documento: son adicionales.

En efecto, los metadatos fueron creados para describir características inherentes a un documento; comenzaron como descriptores básicos del contenido de ese documento, pero después se fueron sofisticando para describir muchos otros atributos del mismo, y también se fueron especializando para describir características específicas de cada tipo de documento. De todos los tipos de documentos, pueden extraerse varios atributos generales, pero además puede extraerse una serie de características o atributos propios y específicos para cada uno de ellos; esas características propias son importantes para el sector que los produce y los utiliza. Además, el peso específico y la importancia de cada una de esas características varía de sector en sector y, por tanto, su tratamiento es muy distinto para cada tipo de documento: libro, archivo, artículo, filme, imagen, partitura, noticia, etcétera. Además, el advenimiento de los documentos digitales trajo nuevas listas de atributos a incluir. En términos generales y agrupando en una lista-resumen, al comparar y extraer diversas especificaciones y conceptos para cada tipo de documento, las características específicas inherentes a todos los tipos de documentos digitales pueden describirse en los siguientes grandes grupos:

- Contenido - Tipos y géneros de contenidos, multimedialidad (variedad de tipos de datos en un mismo documento), multidimensionalidad (múltiples aproximaciones o enfoques a un documento), etcétera.
- Contexto - Información acerca de cómo, cuándo, dónde, por quién y en qué circunstancias se produce el documento; por qué se produce; para qué se usa y su relación con otros documentos en un corpus.
- Proveniencia - Información acerca del origen del documento: persona(s) o autoría, organización, procedimiento, etcétera.
- Estructura - Información acerca de cómo debe ser la apariencia del documento, cómo se integran intelectualmente las partes de documentos complejos, hipervínculos, etcétera.
- Almacenamiento - Información sobre las características particulares del almacenamiento actual del documento, o cómo debe almacenarse (dispositivos especiales, algoritmos de compresión, etcétera).
- Accesibilidad - Medida en que el documento puede ser accedido, reproducido, copiado, modificado, transmitido, etcétera; su gratuidad o costo, etcétera.
- Preservación - Información acerca de cómo se debe preservar o se ha preservado ese documento; periodos de conservación, eliminación, etcétera.
- Autenticidad - Información acerca de los atributos que el documento tiene y debe conservar a lo largo del tiempo para ser considerado auténtico u original (autoría, identidad, integridad, estabilidad, custodia, firmas, etcétera).



- Modificabilidad - Información acerca del grado en que el documento puede ser modificado, en qué condiciones y por quién (wikis, folksonomías, etcétera).
- Tecnología - Datos acerca del entorno tecnológico de producción y uso del documento: sistemas, equipos, y programas informáticos; formatos, estándares, escalas, procesos técnicos efectuados, envolturas, firmas o sellos digitales, sumas verificadoras, cifrados, certificados, etcétera.
- Modularidad - Medida en que el documento puede ser dividido en varias partes para su uso o interpretación.
- Interoperatividad - Medida del grado en que el documento puede ser utilizado a lo largo de múltiples plataformas y estándares con pocas o nulas restricciones.
- Dinamismo - Medida en que el documento se actualiza y cambia con las experiencias del usuario (bases de datos, páginas web dinámicas, etcétera).

No todas las anteriores características existen en todos y cada uno de los documentos digitales, pero absolutamente todos los tipos de documentos digitales tienen varias de estas características. Éstas deben ser representadas por medio de metadatos específicos para cada una de ellas.

Además del contenido, algunos de estos atributos se consignan en forma inherente dentro del propio documento, como el nombre o identificador del archivo informático que lo contiene o el formato principal del documento: doc, pdf, jpg, wav, mp3, etcétera, y muchos otros atributos se agregan en forma de metadatos específicos. Para cada tipo de documento, se contemplan o recomiendan metadatos específicos

con mayor o menor énfasis en algunos de ellos dependiendo de la importancia que se les otorga dentro de ese tipo de documentos. Así, se encuentran metadatos específicos descriptivos, documentales, estructurales, de contexto, jurídicos, administrativos, funcionales, tecnológicos, de autenticidad, de ubicación, de privilegios o restricciones al acceso y uso, de procedimiento, y por supuesto de preservación, etcétera. En función de sus características y volumen, los metadatos pueden estar embebidos dentro del documento o crearse en documentos aparte con sus correspondientes hipervínculos. Más aún, en algunos tipos de documentos se recomienda ya el uso de “metametadatos”; esto es, datos acerca del origen de los metadatos y su compilación, con el propósito de establecer su rigor, precisión, autenticidad, amplitud, etcétera, y por consiguiente, su confiabilidad; por ejemplo, nombre de personas u organizaciones que compilaron los metadatos; fechas de los mismos; si son metadatos generados en forma manual o automatizada; las normas y estándares, fuentes, herramientas, técnicas, procedimientos, actualizaciones, etcétera, utilizados para recopilarlos.

Para abundar más acerca del tema de la variedad de metadatos, consúltese el diccionario PREMIS (2016), así como la norma ISO 23081-1 (2011) - Información y documentación - Procesos para la gestión de documentos – Metadatos. Ambos son de índole general; para cada tipo de documentos existen estándares, normas y recomendaciones específicos. Éstos son hoy en día muy numerosos. A manera de ejemplos pueden mencionarse:

- Los estándares ISO/IEC 15489:2001 - “Información y Documentación - Gestión de Documentos - Parte 1. Generalidades”; ISO/IEC 15489:2006

- “Información y Documentación - Gestión de documentos - Parte 2: Directrices”.
- El estándar ISO TR 15801:2009 -Recomendaciones para la veracidad y confiabilidad de información almacenada electrónicamente.
- El estándar ISO 18492:2005- Conservación a largo plazo de información basada en documentos.
- ISO 19005-1:2008 – Partes 1, 2 y 3. Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1).
- El Modelo de Referencia OAIS –Open Access Information System– para archivos digitales, el cual generó la norma ISO 14721:2012 (OAIS, 2012).
- El Modelo MoReq: Modelo de Requisitos para la gestión de Documentos de Archivo - (DLM Forum, 2001).
- El documento “*Trusted Digital Repositories*” del RLG-OCLC Working Group on Digital Archives Attributes - (RLG-OCLC, 2002).
- La iniciativa NDIIPP - The National Digital Information Infrastructure and Preservation Program de la Biblioteca del Congreso de los E.U.A. - (NDIIPP, 2012).

Como ya se mencionó, el gran problema de la preservación de recursos documentales digitales a largo plazo no reside mayormente en factores de tipo técnico, sino de procedimiento. Independientemente del tipo o modelo de plan de preservación que una organización adopte y del tipo de documentos que maneje, es necesario establecer de entrada una política de preservación, la cual será acompañada de estrategias de *conservación y estrategias de preservación*

que deriven en un plan de acción. Esto obedece de origen a planteamientos ampliamente difundidos por la UNESCO en sus *Directrices para la Preservación del Patrimonio Digital*, que ofrecen un amplio marco de referencia para el establecimiento de estrategias de preservación para cada organización y sus tipos de documentos de archivo digitales (UNESCO 2003).

Estrategias de conservación (también llamadas estrategias de mantenimiento). Consiste en un conjunto coherente de objetivos y métodos para proteger y mantener la accesibilidad de los soportes de los documentos digitales a lo largo de las primeras etapas de una cadena de custodia (InterPARES 2010). En general, consisten en:

- 1) Clara asignación de responsabilidades dentro de la organización.
- 2) Adquirir la infraestructura técnica adecuada.
- 3) Mantenimiento, actualización y cambio de sistemas informáticos.
- 4) Transferencia de datos periódicamente hacia soportes de almacenamiento nuevos y vigentes.
- 5) Adopción de condiciones físicas apropiadas para el almacenamiento de los soportes.
- 6) Redundancia de medios en distintas ubicaciones físicas.
- 7) Plan de seguridad informática en los sistemas.

Estrategias de preservación. Además de las anteriores estrategias de conservación, todo preservador de documentos digitales es responsable de establecer también estrategias de preservación. Consisten en un conjunto coherente de objetivos y métodos para proteger y mantener la usabilidad y estabilidad tecnológica de los contenidos de documentos

digitales a lo largo de todas las etapas de una cadena de custodia (InterPARES 2010). En general, éstas son:

- 1) Preferir estándares abiertos y permanentes (pdf 1-A, odt, tiff, etcétera).
- 2) Usar formatos autodescriptivos como XML.
- 3) Restringir el número de formatos que pueden usarse en la organización; es decir, normalizar en lo posible estándares, estructuras y formatos.
- 4) Establecer e incluir en los documentos metadatos adecuados para el tipo de documentos en cuestión, sin olvidar los de preservación.
- 5) Migrar, refrescar, emular, etcétera (evitar obsolescencia tecnológica).

Como conclusiones, puede establecerse que la preservación de documentos u objetos digitales no consiste en el simple almacenamiento protegido de documentos en un soporte electrónico. Es toda una escuela de pensamiento cuyo objetivo central pretende hacer permanentes esos documentos digitales en el largo plazo —idealmente para siempre—; pretende mantenerlos accesibles y usables; auténticos, íntegros y fiables. La base de toda su teoría implica, además de la tecnología, la organización y observación cuidadosa de políticas y procedimientos de gestión documental que involucren planes y estrategias para una correcta y permanente preservación documental digital. Más aún, debe ser entendida no solo como un método para mantener cadenas de *bits* legibles y utilizables en un futuro, sino como todo un proceso cultural y social que interactúa en forma planeada y armónica con una completa variedad de necesidades y responsabilidades de organizaciones públicas y privadas, y que sirve para apoyar sus actividades cotidianas, su rendición de cuentas

y su memoria institucional como parte de una historia colectiva de una sociedad.

El proceso de preservación de documentos digitales para su resguardo y uso a largo plazo es un reto que implica amplios conocimientos tanto teóricos como metodológicos; decisiones delicadas, fuertes inversiones en recursos humanos calificados y algunas en tecnología; implicaciones legales, obsolescencia de dispositivos, reconversiones de tiempo en tiempo, y ciertos riesgos. Muchos responsables de colecciones, acervos, repositorios o archivos digitales han considerado la posibilidad de preservar o digitalizar algunos de sus fondos documentales pero han pospuesto la decisión de emprender estos proyectos confiando en que llegaría un momento en el que la tecnología y los costos se volviesen estables y competitivos, y los procesos fueran más simples, solo para darse cuenta de que a pesar de que la tecnología y sus costos mejoran cada día, la cantidad y variedad de materiales a digitalizar crece en forma exponencial y cancela de golpe estas ventajas.

Por lo mismo, no es recomendable esperar a que los parámetros se estabilicen totalmente y los riesgos desaparezcan; eso no está sucediendo y no pareciera ser así en un futuro cercano. Por otra parte, la demanda social y la legislación presionan cada día más para que las organizaciones públicas tengan colecciones, repositorios o archivos digitales dinámicos, actuales, funcionales y operativos, además de ser accesibles en forma remota por medios electrónicos. Es imperativo realizar estas actividades, y por ello será mejor estudiar a fondo el problema, entender sus fundamentos, características, ventajas y desventajas; prepararse bien y realizar un plan meticuloso y detallado para el desarrollo de una estrategia específica para estas actividades a fin de evitar cometer fallas al hacerlo bien desde la primera vez,

sin incurrir en costos o riesgos innecesarios, y sin errores que a la larga se deban lamentar y corregir. Como ha podido observarse a lo largo de este texto, los metadatos de preservación forman parte importante de ello.

BIBLIOGRAFÍA

- DLM Forum. 2001. *Moreq -Modelo de Requisitos Para la Gestión de Documentos Electrónicos de Archivo. Véase también: Moreq2 - Modelo de Requisitos Para la Gestión de Documentos Electrónicos de Archivo, versión 2, 2004. Y también: MoReq 2010 - Modelo de Requisitos Para la Gestión de Documentos Electrónicos de Archivo, versión 2010.* Disponible en [http://ec.europa.eu/archival-policy/moreq/index\\_en.htm](http://ec.europa.eu/archival-policy/moreq/index_en.htm).
- InterPARES. 2010. (The International Research on Permanent Authentic Records in Electronic Systems). *Guía del Preservador - Preservación de Documentos de Archivo Digitales: Guía Para las Organizaciones.* México: UNAM - CUIB. Disponible en <http://iibi.unam.mx/archivistica/guias>.
- InterPARES 2. 2019. Glossary Database. Entrada por “Digital Object”. Disponible el 24 de octubre de 2019 en [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm).
- ISO 23081-1. 2011. Información y documentación. *Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1 - Principios; Parte 2 - problemas conceptuales y de implementación; parte 3 - Métodos de autoevaluación.*
- NDIIPP. 2012. (The National Digital Information Infrastructure and Preservation Program of the Library of Congress). *Preserving our digital heritage.* Disponible en [http://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report\\_Post.pdf](http://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report_Post.pdf)



- OAIS. 2012. (Reference Model for an Open Archival Information System). Version CCSDS 650.0-M-2. Management Council of the Consultative Committee for Space Data Systems (CCSDS). Washington DC, junio 2012. Disponible en <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- PREMIS Data Dictionary for Preservation Metadata, Version 3.0*. 2016. Library of Congress. Disponible en <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.
- RLG-OCLC Working Group. 2002. *Trusted Digital Repositories: Attributes and Responsibilities*. An RLG-OCLC Report. Mountain View, CA: OCLC. Disponible en <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.
- UNESCO. 2003. *Directrices para la Preservación del Patrimonio Digital*. División de la Sociedad de la Información de la UNESCO. Preparadas por la Biblioteca Nacional de Australia. Sitio Oficial de la UNESCO. Disponible en <http://www.unesco.org/new/es/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-the-preservation-of-digital-heritage/>.
- Voutssás M., Juan. 2017. *Confianza e Información Digital: Bibliotecas, Archivos y Web*. México: IIBI-UNAM. ISBN: 978-607-02-9757-1. Disponible en [http://132.248.242.6/~publica/archivos/libros/confianza\\_informacion\\_digital\\_s.pdf](http://132.248.242.6/~publica/archivos/libros/confianza_informacion_digital_s.pdf).
- Voutssás-M., Juan y Alicia Barnard. 2014. *Glosario de preservación archivística versión 4.0*. México: IIBI-UNAM. ISBN: 978-607-02-5744-5. Disponible en [http://132.248.242.6/~publica/archivos/libros/glosario\\_preservacion\\_archivistica\\_digital\\_v4.0.pdf](http://132.248.242.6/~publica/archivos/libros/glosario_preservacion_archivistica_digital_v4.0.pdf).

Voutssás-M, Juan. 2009. *Preservación del Patrimonio Documental Digital en México*. México: CUIB-UNAM. ISBN: 978-607-02-0583-5. Disponible en [http://132.248.242.6/~publica/archivos/libros/preservacion\\_patrimonio.pdf](http://132.248.242.6/~publica/archivos/libros/preservacion_patrimonio.pdf).