

TIES Revista de Tecnología e Innovación en Educación Superior

Publicación semestral • ISSN 2683-2968 • Febrero 2024 • Número de revista 9
doi.org/10.22201/dgtic.26832968e.2024.9

Preservación Digital

Editorial

Juan Voutssás Márquez

Derivaciones del uso de la Inteligencia Artificial (IA) en la preservación digital

Perla Olivia Rodríguez Reséndiz

Análisis del estado actual del almacenamiento masivo para la preservación digital

José Antonio Contreras Benítez, Ma. Guadalupe Sánchez Mendoza, Maribel González González, Omar Solís Garza

Criterios para valorar Sistemas de Preservación Digital

Ana Yuri Ramírez Molina, Carlos Alberto Román Zamitiz, Ivan Pérez Pineda, José Antonio Contreras Benítez
José Antonio Salazar Carmona, Juan Manuel Castillejos Reyes, Maribel González González, Teresa del Rocío González Melchor

Preservación Digital y Derechos de Autor, una visión analógica en un mundo digital:
problemas de regulación en la Ley de Derechos de Autor en España

Miguel Ángel Corado Guerrero

La Hemeroteca Nacional Digital de México:
perspectivas de la preservación digital para el patrimonio documental

Alberto Castro Thompson, Lisandro Pablo Olivares

Preservación y acceso a objetos digitales fuera de línea: experiencias en la Filmoteca de la UNAM

Gerardo León Lastra, Gustavo Lucio José, Manuel Comi Xolot, Luis Felipe Maciel Mercado



EDITORIAL

JUAN VOUTSSÁS MÁRQUEZ

*“Las viejas cadenas de bits nunca mueren;
solo se vuelven ilegibles”*

*Jeff Rothenberg, científico senior de la
Corporación Rand, 1995.*

En el año 2005, siendo yo Director de Sistemas en la Dirección General de Servicios de Cómputo Académico de la UNAM, llegó el tiempo de apagar la computadora A-12 de Unisys, debido a su obsolescencia. Esta máquina era el último de los mainframes o grandes computadores centrales que quedaba en la institución de entre toda una larga serie de este tipo de equipos utilizados en esta casa de estudios desde que en ella se inició el cómputo en Iberoamérica en 1958.

Debido a su edad y tipo esa máquina –entre sus dispositivos de entrada-salida– contaba todavía con una unidad lecto-grabadora de cintas magnéticas de carrete abierto: esas que aparecen en los filmes como un gran anaquel en el cual se observan dos grandes bobinas girando simultáneamente. La función de estos carretes era almacenar en ellos información en forma magnética, aquella que no se deseaba tener en línea en los discos duros, o como respaldo de la misma. Ese tipo de unidades habían entrado en servicio con la mismísima Univac I –la primera computadora comercial puesta a la venta en 1951– y desde entonces, con algunas pequeñas mejoras, habían sido el medio idóneo para almacenar información electrónica a lo largo de casi cinco décadas.

El advenimiento del cd-rom hacia fines de los ochenta, seguido del DVD una década después, así como las cintas magnéticas ya insertadas dentro de un cartucho, cambiaron ese panorama, abaratando sensiblemente el almacenamiento y llevándolo a otras magnitudes mucho mayores. Las clásicas unidades de carrete abierto se volvieron obsoletas: por más de una década nadie las compraba más, y por lo mismo nadie las fabricaba. Los centros de cómputo de la UNAM habían mudado ya desde hacía largo rato todo su acervo de cintas hacia otros medios más modernos.

Poco antes de apagar la A-12, a sabiendas de que era la última en la universidad con esa unidad lecto-

ra, se decidió ofrecer un servicio a la comunidad universitaria para convertir esas cintas hacia otro dispositivo más moderno, pensando que “tal vez” alguien poseyese una de ellas todavía. Al efecto se hizo un anuncio en Gaceta UNAM. ¡Cuál sería mi sorpresa al ver que una gran cantidad de personas comenzaron a aparecer con sus cintas de carrete abierto! El “tal vez” se volvió un desfile cotidiano, y una vez que se corrió la voz, cada vez más personas acudían a solicitar ese servicio, gratuito por cierto. Hubo quienes inclusive llegaron con un “diablito” lleno de cintas para requerirlo. Lo que pensábamos que iba a ser el caso aislado de unos pocos universitarios que guardaban alguna como recuerdo se convirtió en un proceso de cientos de personas a lo largo de casi tres meses. El servicio fue incluso requerido por personas e instituciones fuera de la UNAM. Muchos aprovecharon esta última oportunidad de actualizar su información conscientes que de otra forma se hubiera perdido.

Este caso ilustra con claridad el problema de la “obsolescencia tecnológica”, especialmente el de la información. Es indiscutible que el auge e incremento de la información digital en las últimas décadas trajo consigo un sinnúmero de beneficios, y hasta cambió nuestra forma de percibir el mundo. Hoy en día una enorme proporción de la humanidad genera y consume información digital, y por lo mismo, es necesario guardarla. Pero al momento mismo en que cada pieza de información se guarda, el fantasma de la obsolescencia aparece para amenazarla: y no son solo los dispositivos, eso es tal vez lo más notorio; también están los programas o aplicaciones que leen una pieza de información y que requieren además del uso de un formato específico para hacerlo; y esos programas corren bajo un cierto sistema operativo. Si no se posee el dispositivo lector, con el programa adecuado, dentro de un computador que tenga el sistema operativo al efecto, el proceso de leer un fichero almacenado en un cierto soporte y formato se vuelve imposible, y la pieza de información ahí contenida, al ser ilegible, no está ya disponible y por tanto se ha perdido. Una pregunta que viene al caso: ¿Usted posee disquetes con cierta información? ¿Cuántos de ellos puede todavía leer?

Cuando se habla de preservar un objeto físico: un libro, un manuscrito, una fotografía, un disco musical, etcétera, se trata de mantenerlo lo más parecido a su condición original; que nada lo altere, que nada lo cambie, así el objeto se mantendrá por largo tiempo: se ha preservado. En el mundo digital sucede lo opuesto: si nadie lo toca, si nada lo cambia, si nunca se altera, al cabo de un cierto tiempo el objeto contenido, a pesar de que sigue existiendo, será inaccesible, debido precisamente a la obsolescencia tecnológica. De aquí se desprende el paradigma fundamental de la preservación digital: para preservar una cierta información de este tipo, es indispensable cambiarla de tiempo en tiempo, una y otra vez. No obstante, se debe mantener el contenido inalterado y, a pesar del cambio, se debe poder garantizar su autenticidad e integridad, además de su existencia y acceso.

En un principio se intentó abordar el problema tratando de extender en lo posible la vida de los soportes físicos: discos y cintas de larga duración, ambientes controlados sin oxígeno y a temperaturas ideales, etcétera. Pronto se vio que –aunque eso en efecto extendía la vida de esos soportes– igualmente los formatos, los programas y los sistemas operativos se hacían obsoletos, y de todas formas el contenido se hacía inaccesible de forma muy rápida.

Se entendió entonces que la preservación digital no era un problema que dependiera de los soportes de la información. Siendo estos importantes, lo más relevante consistía en los métodos para preservarla. El enfoque cambió entonces, y los esfuerzos se concentraron en los procesos, en las técnicas, en los estándares para lograrlo. En la actualidad está totalmente aceptado que si bien el problema de la obsolescencia digital proviene del uso de la tecnología, no es en ella en donde reside su solución. Preservar información digital no consiste en comprar equipos y dispositivos para ese propósito, sino en implantar en cada institución toda una metodología al efecto. De aquí se desprende que preservar información digital es un proceso que se instaura, no un producto que se compra. Una vez que se comprende el paradigma fundamental de la preservación digital y se comienza la instauración de un proceso se está en el camino correcto para ello.

Pero instaurar un proceso implica el desarrollo de un plan, con políticas de preservación, con procedimientos, estándares, buenas prácticas, personal capacitado, y sí, al final, con computadores y programas al

efecto. La preservación digital es 80 por ciento método y 20 por ciento equipo y sistemas.

De todo lo anterior, cabe resaltar que la preservación digital es un proceso que busca que los contenidos digitales trasciendan a sus soportes, manteniendo esos contenidos estables, accesibles y auténticos a lo largo de múltiples generaciones tecnológicas; es por tanto un proceso ininterrumpido y a largo plazo.

Por tanto, si la preservación digital consiste mayormente en método, en un proceso, la única manera de implantarlo en una cierta institución es por medio de personal capacitado al efecto, que entienda perfectamente que preservar lo digital va mucho más allá de solo obtener unas simples copias de respaldo de la información: implica una cultura de la preservación, un conocimiento profundo de las características de la información propia de cada institución, un entendimiento adecuado de las técnicas, estándares, metodologías y recomendaciones para lograrlo. Por ello es de especial importancia que ese personal se adentre y profundice dentro de todo ese saber.

El presente número de la revista está dedicado expresamente al tema de la preservación digital, y por lo mismo se conforma como un compendio que contiene varios ejemplos donde cada uno de ellos ilustra de manera muy clara y detallada diferentes aspectos de lo aquí mencionado: conocimiento, técnicas, estándares, buenas prácticas, métodos, etcétera, precisamente para con todo ello hacer evidente la amplitud del tema y ayudar a otros interesados a profundizar en el mismo. Pueden verse dos textos acerca de almacenamiento para preservación, tanto en línea como fuera de línea; uno de criterios para la evaluación de sistemas de preservación; otro que analiza sus aspectos legales; uno más que hace una reseña de las perspectivas de la preservación en una hemeroteca digital, y finalmente uno que reflexiona acerca de los usos de la Inteligencia Artificial en la preservación. Cada uno de los artículos destaca aspectos importantes de entre todos los posibles a considerar; a través de ellos puede el lector darse cuenta que la preservación digital no es algo deseable sino indispensable, no es algo trivial, no consiste en simples respaldos o copias de la información, no es algo eventual o casual, y puede con todo el conjunto aprender mucho de los métodos al respecto.

La Dra. Luciana Duranti, una experta en el tema a lo largo de varias décadas, expresó una vez que la esencia de la preservación digital consistía en entender que

en realidad, no intentamos preservar documentos digitales, sino nuestra capacidad de accederlos, incólumes, una y otra vez a través de largo tiempo.

Conviene pues, entender esa esencia y aprender esos métodos.



TIES Revista de Tecnología e Innovación en Educación Superior

DERIVACIONES DEL USO DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA PRESERVACIÓN DIGITAL

DOI: 10.22201/dgtic.26832968e.2024.9.1

Perla Olivia Rodríguez Reséndiz (perla@iibi.unam.mx)
*Universidad Nacional Autónoma de México, Instituto de Investigaciones
Bibliotecológicas y de la Información. Ciudad de México, México.*
ORCID: 0000-0001-6517-8895

www.ties.unam.mx

Fecha de recepción: marzo de 2023 • Fecha de publicación: febrero de 2024

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

DERIVACIONES DEL USO DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA PRESERVACIÓN DIGITAL

Resumen

La capacidad de generar datos es constante e imparable. Cada año producimos más información que el anterior y su preservación digital es un problema social que afrontan las instituciones de la memoria, entre las que destacan los archivos sonoros y audiovisuales. Ante esta situación, la gestión y el manejo de grandes volúmenes de información digital a través de la Inteligencia Artificial (IA) podría significar una alternativa, cuyas derivaciones deben ser analizadas desde diferentes perspectivas.

En este artículo se comparte una reflexión en torno al empleo de la IA en la preservación digital, cuya incorporación resulta paradójica. La IA ofrece oportunidades de mejora en los procesos documentales y, al mismo tiempo, podría derivar en consecuencias adversas. Para establecer las derivaciones de la IA en los archivos sonoros y audiovisuales, se realizó una investigación cualitativa. Como resultado, se advierten costos medioambientales, patrimoniales y sociales, cuyo impacto en la preservación digital debe ser considerado, a fin de proponer estrategias de mitigación.

Palabras clave:

Inteligencia artificial, archivo, datos masivos, preservación digital.

ARTIFICIAL INTELLIGENCE (AI) DERIVATIONS IN DIGITAL PRESERVATION

Abstract

The generation capacity of data is constant and unstoppable. Each year we produce more information than the previous one and its digital preservation is a social problem facing sound and audiovisual archives, among other heritage institutions. In this situation, the management and handling of large volumes of digital information through Artificial Intelligence (AI) could represent an alternative whose derivations should be analyzed from different perspectives.

This article reflects on the use of AI in digital preservation, the incorporation of which seems paradoxical. It offers opportunities for improvement in documentary processes and at the same time could have adverse consequences. Qualitative research was carried out to establish the implications of AI in sound and audiovisual archives. As a result, environmental, patrimonial and social costs were identified, whose impact on digital preservation should be considered in order to suggest mitigation solutions.

Keywords:

Artificial intelligence, archiving, massive data, digital preservation.

DERIVACIONES DEL USO DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA PRESERVACIÓN DIGITAL

1. Introducción

La inteligencia artificial (IA) se ha insertado en nuestra vida cotidiana como una herramienta para el procesamiento de grandes volúmenes de datos. Por medio de la IA se identifica, clasifica, predice y recupera información registrada en una amplia gama de formatos y lenguajes (textos, imágenes y sonidos). Se augura que el desarrollo científico se verá beneficiado, por ejemplo, en la salud, la biología y la genética para el tratamiento de enfermedades. Por otra parte, su inserción en la industria de los datos proyecta un crecimiento continuo en la economía mundial [1].

En fechas recientes, la IA ha devenido en una tecnología popular y polémica. Millones de personas utilizan programas como *Chat GPT (Open AI)*, *Bard (Google)* y *Bing (Microsoft)*, para realizar, desde una tarea escolar, hasta un discurso político. En contraste con la exaltación social que produce el uso de la IA, se han formulado preocupaciones y visiones críticas en torno a que podría degradar la ciencia y empobrecer la ética por ofrecer una visión errónea del lenguaje y del conocimiento [2].

Además, se ha señalado que la IA genera información a partir de datos publicados, sin hacer referencia a la propiedad intelectual, incurriendo con ello en una forma de plagio. Se alerta que puede ser una herramienta para la manipulación, a través de la generación de información falsa y *fake news*. Por ello, cada vez más foros y publicaciones ofrecen disertaciones sobre las implicaciones éticas del uso de la IA.

En este trabajo se abordará la relación de la IA con las instituciones de la memoria, es decir, los archivos, bibliotecas y museos, cuya misión es salvaguardar el patrimonio documental. De estas, se decantó por los archivos, donde se preservan diferentes tipos de documentos, como los sonoros y los audiovisuales, que salvaguardan importantes volúmenes de información, digitalizada o de origen digital, en los cuales la incorporación de la IA se percibe como una herramienta potencial que puede contribuir a la preservación digital, aunque su aplicación no está exenta de posibles efectos adversos.

El objetivo de este artículo es identificar las derivaciones asociadas con el uso de la inteligencia artificial en la preservación digital. Para ello, se optó por el empleo de la metodología cualitativa. Se presentan los antecedentes del uso de la IA en los archivos sonoros y audiovisuales. Como ejemplo, se abordan las primeras experiencias de investigación aplicada en el uso de IA en este tipo de instituciones de la memoria. Además, se definen y relacionan las nociones preservación digital, datos masivos e inteligencia artificial. Finalmente, se formulan algunas de las posibles implicaciones o los costos de carácter medioambiental, patrimonial y social, derivados del uso de la IA en la preservación digital.

2. La IA en la preservación digital

Los archivos se han erigido en espacios vinculados al poder político y al conocimiento de toda sociedad.

Preservan el pensamiento, las creaciones y los hechos que dan cuenta del acontecer de la sociedad. En la era de la información digital, los archivos han adquirido notoriedad por el acceso abierto y la transparencia en la información pública, cualidades esenciales de toda sociedad democrática.

Aun cuando en muchos casos subsiste el rezago en la digitalización de las colecciones analógicas, cada día se incrementa el volumen de documentos de origen digital, cuya preservación es una tarea inaplazable y compleja. Aplicar procesos documentales a grandes volúmenes de información es una tarea titánica. Por ello, desde hace más de una década, la IA ha despertado el interés de los archivistas para mejorar procesos como la catalogación, así como para recuperar contenidos de manera más eficiente.

De todos los tipos de archivos, destacan los que preservan contenidos sonoros y audiovisuales, con los cuales se han iniciado proyectos que ponen a prueba las posibilidades de la IA. Un ejemplo de ello es la experiencia del *American Archive of Public Broadcasting (AAPB)* de Estados Unidos, donde se han probado las posibilidades de usar el aprendizaje automático, en inglés *Machine Learning (ML)*, y la inteligencia artificial [3]. Otro caso es la iniciativa de Radio y Televisión Española (*RTVE*) para la generación automática de metadatos, a través de la transcripción automática de texto y segmentación de contenidos a fin de dar visibilidad a los materiales de archivo [4]. Asimismo, conviene destacar el desarrollo de *AIA Media* consorcio de investigación financiado por la Unión Europea y donde participan universidades, centros de investigación, archivos y medios públicos.

El primer intento por utilizar IA en los archivos sonoros y audiovisuales en México inició hace una década. La doctora Mireya Saraí García, del Centro de Investigación y Desarrollo de Tecnología Digital del Instituto Politécnico Nacional, lideró *MEX CULTURE*, un proyecto de investigación aplicada, encaminado a probar herramientas de inteligencia artificial para la recuperación de contenidos guardados en archivos sonoros [5]. Este propósito fue complejo porque hace una década apenas habían comenzado las tareas de digitalización. Por ello, la preservación de documentos de origen digital era una práctica desconocida. No había suficientes contenidos para aplicar y probar la IA en México. Además, para los responsables de los archivos la idea de aplicar algoritmos en una copia de los contenidos digitales era incomprensible y suponía un riesgo.

Así, las primeras aplicaciones de inteligencia artificial en archivos, promovidas por investigadores mexicanos,

florecieron en Colombia. En Radio y Televisión Nacional de Colombia se creó el Laboratorio de Inteligencia Artificial de RTVC, impulsado por la maestra Dora Braussin. Desde entonces esta iniciativa es un referente en América Latina.

En 2019, el Instituto de Investigaciones Bibliotecológicas y de Información de la UNAM, en colaboración con la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, organizó el Congreso Internacional de Archivos Digitales, con el tema *Inteligencia Artificial y Big Data*. En este Congreso se presentaron algunas de las experiencias más relevantes de Europa y Estados Unidos en la aplicación de la IA en archivos audiovisuales.

Por la relevancia del tema, se emprendieron investigaciones en el ámbito de las Ciencias Sociales y las Humanidades en la UNAM. Entre otras, conviene señalar que inició la primera tesis de doctorado sobre el uso de la IA en archivos sonoros, a cargo de la maestra Georgina Sanabria Medina. En 2020 se creó, por iniciativa del doctor Pedro Salazar Ugarte, la Línea de Investigación en Derecho e Inteligencia Artificial (*LIDIA*) del Instituto de Investigaciones Jurídicas de la UNAM. En marzo de 2023, *LIDIA* y el Instituto Nacional de Acceso a la Información (*INAI*), organizaron el foro “Retos del uso de Inteligencia Artificial en la actividad archivística de México”.

En la actualidad, la IA y los archivos son un ámbito de interés científico y profesional.

3. Las derivaciones

Para garantizar la permanencia de los objetos digitales y los metadatos, la preservación digital no debe ser interrumpida. Los datos se almacenan en archivos o centros de datos, a través de los cuales se tiene acceso a la información de un determinado ámbito o negocio. Un ejemplo de ello son los servicios de información que ofrecen *Alexa* y *Siri* como asistentes virtuales. Cuando se formulan consultas, entra en operación un complejo y vasto sistema de aplicaciones informáticas, que procesan, en cuestión de segundos, millones de datos para arrojar de manera inmediata la información solicitada. En este proceso intervienen dos términos: el *big data* y la inteligencia artificial. El *Big data* procesa y recupera grandes volúmenes de información, gracias al uso de la inteligencia artificial [6]. El empleo de la IA en el procesamiento de grandes volúmenes de datos para la preservación digital, tiene al menos tres consecuencias o costos: medioambiental, patrimonial y social, que a continuación se desarrollan.

4. Medioambiental

Los centros de datos son instalaciones para la conservación y la gestión de sistemas informáticos y componentes asociados, como telecomunicaciones y sistemas de almacenamiento. Cuentan con un entorno seguro y una temperatura controlada [7]. En concordancia con la definición anterior, los archivos digitales constituyen un tipo de centro de datos.

Los centros de datos son parte de la sociedad contemporánea, porque preservan grandes volúmenes de información digital, tarea para la cual se requiere del uso de energía eléctrica de manera ininterrumpida, no sólo para la conservación a través del almacenamiento de los datos, sino también para llevar a cabo los procesos de búsqueda y recuperación de información. En otros términos, que la búsqueda de información desde cualquier dispositivo conectado a Internet genera un impacto en el medio ambiente.

Los centros de datos dejan una huella en el medio ambiente porque consumen energía y agua constantemente. Utilizan entre 10 y 50 veces más energía que una oficina. En 2021, emplearon el 1% de la demanda mundial de energía y millones de galones de agua para enfriar los sistemas de almacenamiento y generar energía [7].

Estudios recientes señalan que, para 2040, el sector de las tecnologías de la información emitirá el 14% de las emisiones de carbono y que la demanda de energía en los centros de datos se multiplicará por 15 en 2030 [8]. Este incremento está asociado al uso de la inteligencia artificial, porque para el procesamiento de grandes volúmenes de datos se emplean arquitecturas basadas en redes neuronales, con millones de parámetros que se deben entrenar [8].

Las empresas que mantienen los grandes centros de datos no proporcionan información sobre la huella de carbono, el uso de agua para enfriar los sistemas de almacenamiento y la producción de basura tecnológica que se genera al entrar en desuso la tecnología cada determinado tiempo [9]. Diseñan planes para reducir el impacto en el medio ambiente y afirman que contribuirán a que la información proporcionada, mediante los centros de datos, ayude a disminuir las emisiones mundiales de gases de efecto invernadero, entre un 6 y 12%, en 2030 [7].

No obstante, las estimaciones actuales proyectan que el impacto ambiental de las tecnologías de la información se acrecentaría con el uso de la IA, al ser una herramienta utilizada por millones de personas en el

mundo y porque, precisamente, durante el proceso de entrenamiento para perfeccionar su desempeño, requiere de mejores procesadores de información [8].

5. Patrimonial

Para garantizar la preservación de grandes volúmenes de datos, se requieren sistemas robustos de almacenamiento digital que emplean la nube como una solución. La preservación en la nube es una decisión de los responsables de los archivos, que tiene implicaciones patrimoniales y se relaciona con la noción de la soberanía digital.

El cómputo en la nube se refiere al uso de servidores, ubicados en lugares remotos, que a través de Internet almacenan, gestionan y procesan datos, en lugar de utilizar un servidor local o personal. Con esta infraestructura, las aplicaciones compartidas a través de Internet pueden ser utilizadas por múltiples usuarios. Los servicios de cómputo en la nube pueden ser proporcionados por proveedores externos, que ofrecen y mantienen el *software* y el *hardware* necesarios [7] para la preservación de grandes volúmenes de información digital.

Esta solución puede ser atractiva porque resuelve de manera práctica la gestión y la conservación de datos en un archivo. Sin embargo, se observa cada vez más resistencia a la dependencia de empresas como *Amazon*, *Google* o *Microsoft* en materia de almacenamiento y gestión de grandes volúmenes de datos. Este tema cobró relevancia en la agenda política de Europa desde 2019 [10].

El Parlamento Europeo reconoció que los datos y la inteligencia artificial son elementos para la innovación y para resolver problemas sociales, desde la salud y la agricultura, hasta la seguridad. Por ello, asignaron 20 millones de euros anuales para el desarrollo de la inteligencia artificial [10]. Con esta determinación se reconoció la importancia de invertir para crear soluciones de almacenamiento digital que no vulneren la soberanía digital. La experiencia europea debería motivar la inclusión del tema en la agenda política de todas las naciones, para sentar las bases de la soberanía digital y garantizar la protección del patrimonio.

6. Social

La manipulación en el procesamiento de grandes volúmenes de datos conlleva riesgos sociales. A los patrones de entrenamiento de la IA se pueden trasladar los prejuicios y las discriminaciones que ya existen en la sociedad, por razón de sexo, género, religión, opiniones políticas,

discapacidad, orientación sexual y edad, entre otras [11]. Esto supone imponer la visión de una persona o de un grupo de personas sobre los modos de interpretación y recuperación de un conjunto de datos.

La preocupación por el uso indebido de la IA motivó a investigadores y tecnólogos a firmar una carta para advertir sobre los riesgos que para la sociedad y la humanidad acarrearía el desarrollo sin control de la IA [12]. Se señala en la misiva que el desarrollo de la tecnología está fuera de control y podría crear mentes digitales cada vez más poderosas, que ni siquiera sus creadores podrían entender o predecir. En consecuencia, se formuló un llamado para detener, al menos seis meses, el desarrollo de la tecnología, a fin de poner en marcha protocolos de seguridad y crear un marco regulatorio en el que se involucren los legisladores de cada país para la supervisión y el seguimiento de los grandes sistemas de información que emplean IA [12].

Las posibilidades de manipulación y uso indebido de la IA en grandes volúmenes de datos dejan al descubierto el costo social. El tratamiento de la información puede incidir en la toma de decisiones de la sociedad e incluso vulnerar la democracia [6].

La discusión de la IA en la preservación digital debe basarse en la transparencia del procesamiento de la información, la mitigación de la discriminación de cualquier tipo y la garantía de que no se vulneren los derechos individuales en torno a la posible toma de decisiones automatizadas. El uso erróneo de la tecnología afectaría la confianza y la credibilidad en el archivo. La IA es una herramienta que puede proteger y crear modos creativos, que faciliten el acceso sencillo y más amplio a los datos de los bienes culturales digitales que forman parte del patrimonio digital [13].

7. Conclusiones

En la preservación digital de grandes volúmenes de datos el uso de IA es paradójico. Podría tener aplicaciones que beneficien los procesos documentales, para asegu-

rar la permanencia, el valor y el acceso a los contenidos, pero al mismo tiempo acarrear consigo costos o implicaciones adversas. En relación con lo cual, se identifican tres posibles derivaciones: medioambiental, patrimonial y social. En referencia a la medioambiental, se advierte que el uso de la IA podría acrecentar, en la preservación digital, el uso de energía eléctrica, grandes cantidades de agua y la generación de basura tecnológica, y con ello incidir en la afectación de las tecnologías de la información en el medio ambiente.

El costo patrimonial alude a que el volumen de datos preservados es tan grande que en algunos casos se emplean soluciones comerciales como la nube. Su utilización podría significar un peligro que vulnere la soberanía y sitúe en condición de riesgo al patrimonio digital. De ahí que sea necesario invertir en tecnología y en el desarrollo de infraestructura nacional para la preservación y almacenamiento digital de grandes volúmenes de datos. Asimismo, es necesario considerar que el tema debe ser parte de la agenda política nacional de cada país. El costo social se refiere a los riesgos asociados al uso indebido y la manipulación de la información. El uso indebido de la IA en el tratamiento documental afecta la credibilidad y la confianza, valores sobre los que un archivo se sustenta. Además, ubica en condición de vulnerabilidad a la sociedad porque pone en peligro los derechos humanos.

La disertación en relación con las derivaciones o costos, medioambiental, patrimonial y social, debe ser parte de la discusión pública. Es necesaria para diseñar y poner en marcha estrategias que aminoren el posible impacto negativo de la IA en la preservación digital. Entre otros fundamentos, se advierte el establecimiento de los éticos en el uso de la IA, que prevengan un uso social inadecuado; la creación de alternativas que aminoren el futuro impacto de la tecnología en el medio ambiente, y la puesta en marcha de estrategias tecnológicas que protejan el patrimonio nacional de los pueblos y no vulneren su soberanía.

BIBLIOGRAFÍA

[1] Comisión Europea. 2020. "Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza" [COM (2020) 65 final]. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

[2] N. Chomsky, I. Roberts y J. Watumull. "Noam Chomsky: The False Promise of ChatGPT". *The New York Times*, 23 de marzo de 2023. En línea: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

[3] K. Cariani, Karen y D. Ives. "Using Computational Tools and Experts to Improve Access to Digital Media Archive". En *Inteligencia artificial y datos masivos en archivos digitales sonoros y audiovisuales*, pp. 101-115. México: UNAM-IIBI, 2020.

[4] V. Bazán Gil. "Reflexiones en torno a la IA: el caso de RTVE". En *Inteligencia artificial y datos masivos en archivos digitales sonoros y audiovisuales*, pp. 117-131. México: UNAM-IIBI, 2020.

[5] M. García Vázquez, Y. Montiel Pérez, J. Benois- Pineau y M. Crucianu, Michel. "Plataforma multimedia Mex-Culture". En *Archivos Digitales Sustentables. Conservación y acceso a las colecciones sonoras y audiovisuales para las sociedades del futuro*, pp. 141-155. IIBI UNAM, 2017.

[6] D. Kusnetzky. "What is Big Data?" *ZDNet*. 2023. Disponible en: <https://www.zdnet.com/article/what-is-big-data/>

[7] Digital Climate Alliance (DCA). "Sustainable data centers: Powering the digital revolution". *Digital Climate Alliance*. 2023. Disponible en: <https://www.digitalclimate.io/data-center-white-paper>

[8] M. Pascual. "El sucio secreto de la inteligencia artificial". *El País*. 2023. Disponible en: <https://elpais.com/tecnologia/2023-03-23/el-sucio-secreto-de-la-inteligencia-artificial.html>

[9] L. Tadic. "La relación entre archivos audiovisuales y el medioambiente". En *Cambio climático y preservación digital sonora y audiovisual*. Quito, EC: Universidad Andina Simón Bolívar, Sede Ecuador / Red Iberoamericana de Preservación Digital de Archivos Sonoros y Audiovisuales, RIPDASA / Ciencia y Tecnología para el Desarrollo, CYTED, pp. 11-28. 2022.

[10] LL. Pellicer e I. Fariza. "Europa ultima un plan para dar la batalla en el negocio de los datos". *El País*, 16 de noviembre de 2019. Disponible en: https://elpais.com/economia/2019/11/16/actualidad/1573926886_318836.html

[11] UK Information Commission Officer. "Guidance on AI and data protection" *ICO, UK*. 2023. Disponible en: <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/>

[12] Institute Future of life. "Pause Giant AI Experiments: An Open Letter", 30 de marzo de 2023. Disponible en: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

[13] Parlamento Europeo. "La inteligencia artificial en los sectores educativo, cultural y audiovisual". *Resolución del Parlamento Europeo*, 19 de mayo de 2021. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_ES.html

Fecha de recepción: 30 de marzo, 2023

Fecha de publicación: febrero, 2024

Cómo se cita

P. O. Rodríguez Reséndiz, "Derivaciones del uso de la Inteligencia Artificial (IA) en la preservación digital", *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero, 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

ANÁLISIS DEL ESTADO ACTUAL DEL ALMACENAMIENTO MASIVO PARA LA PRESERVACIÓN DIGITAL

DOI: 10.22201/dgtic.26832968e.2024.9.2

José Antonio Contreras Benítez (antoniocb7@gmail.com)
Universidad Nacional Autónoma de México, Dirección
General de Personal, Dirección de Sistemas. Ciudad de
México, México.
ORCID: 000-0001-8442-6030

Ma. Guadalupe Sánchez Mendoza (lupita@unam.mx)
Universidad Nacional Autónoma de México, Instituto de
Investigaciones Bibliográficas. Ciudad de México, México.
ORCID: 0009-0005-2221-8865

Maribel González González (pdigital@unam.mx) *Profesional
independiente. Ciudad de México, México.*
ORCID: 0000-0002-1430-7412

Omar A. Solís Garza (omar.solis@dgru.unam.mx)
Universidad Nacional Autónoma de México, Dirección General
de Repositorios Universitarios. Ciudad de México, México.
ORCID: 0009-0008-4452-1576

www.ties.unam.mx

Fecha de recepción: marzo de 2023 • Fecha de publicación: febrero de 2024

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

ANÁLISIS DEL ESTADO ACTUAL DEL ALMACENAMIENTO MASIVO PARA LA PRESERVACIÓN DIGITAL

Resumen

El almacenamiento masivo (AM) es un elemento importante y crítico que debe ser considerado por cualquier institución que desee emprender labores de Preservación Digital (PD). Debido a la falta de información sobre el tema, el Grupo de Almacenamiento Masivo (GAM) definió como alcance inicial documentar y analizar el estado actual de las necesidades básicas del almacenamiento para la PD.

En este contexto y como meta específica, se diseñó, difundió y aplicó un instrumento de preguntas específicas que arrojó información sobre el interés del almacenamiento con enfoque de preservación, el tamaño del acervo, los tipos de contenido y los formatos a preservar, el crecimiento, la infraestructura actual y la requerida, la disponibilidad, los medios, los respaldos y copias, los esquemas de recuperación y los problemas comunes en el almacenamiento, entre otros.

Durante una primera etapa de aplicación, el instrumento tuvo aceptación en 32 instituciones de México y Latinoamérica. Algunas de las entidades participantes están relacionadas con la conservación y la preservación del patrimonio documental.

Una vez analizada la información, se observa que la mayoría de las instituciones tiene interés por la PD por normatividad, más que por convicción. Sólo algunas cuentan con infraestructura especializada, mientras que otras no tienen copia de su acervo y no realizan respaldos. Sin duda, las problemáticas más comunes son la carencia de mantenimiento a la infraestructura, la falta de personal calificado, la poca capacidad de almacenamiento y la obsolescencia.

Por otro lado, se constató que instituciones académicas y de carácter nacional, como la Universidad Nacional Autónoma de México (UNAM), están encaminadas hacia el logro de procesos de almacenamiento masivo, enfocados en la preservación digital, porque tienen infraestructura dedicada, cuentan con tres o más copias distribuidas de su acervo y están provistas de almacenamiento expreso para la preservación, tomando como referentes modelos y estándares internacionales.

Palabras clave:

Almacenamiento masivo, respaldos, infraestructura de datos, preservación digital, resguardo de acervos digitales.

CURRENT STATE ANALYSIS OF MASS STORAGE FOR DIGITAL PRESERVATION

Abstract

Mass storage is an important and critical element that should be considered by any institution that needs to undertake Digital Preservation work. Due to the lack of information on the subject, the Mass Storage Group defined as initial scope; to document and analyze the current state of basic needs of mass storage for Digital Preservation.

In this context and as a specific goal, an instrument was designed, disseminated and applied, consisting of specific questions that provide information on: interest in storage with a focus on preservation, size of the collection, types of content and formats to be preserved, growth, current and required infrastructure, availability, media, backups and copies, recovery schemes and common storage problems, among others.

During the first stage of application, the instrument was used in 32 institutions in Mexico and Latin America. Some of the participating institutions are related to the conservation and preservation of documentary heritage.

After analyzing the information, it can be observed that: most of the institutions reflect that the interest in Digital Preservation is due to regulations rather than conviction, only some have specialized infrastructure, others do not have copies of the collection and do not make backups. Undoubtedly, the most common problems are: infrastructure maintenance, lack of qualified personnel, storage capacity and obsolescence.

On the other hand, it was found that academic and national institutions, such as the Universidad Nacional Autónoma de México, are moving towards achieving mass storage processes focused on Digital Preservation, since they show the existence of dedicated infrastructure, have 3 or more distributed copies of the collection, and designate specific storage for preservation, taking international models and standards as references.

Keywords:

Mass storage, backups, data infrastructure, digital preservation, digital collection storage.

ANÁLISIS DEL ESTADO ACTUAL DEL ALMACENAMIENTO MASIVO PARA LA PRESERVACIÓN DIGITAL

1. Introducción

Dentro de Grupo de Preservación Digital¹, en febrero de 2021 se conforma el de Almacenamiento Masivo (GAM) con el objetivo principal de identificar, documentar, analizar y proponer tecnologías de almacenamiento masivo de información a las instituciones interesadas en la preservación digital (PD).

Considerando que al interior del grupo no se contaba con información sobre la temática en ese momento (datos de almacenamiento de archivos establecidos en el modelo OAIS²; niveles establecidos por la NDSA³ para el componente de almacenamiento; las mejores prácticas interna-

cionales para el almacenamiento en la PD⁴; los sistemas resilientes, y lo que puntualizan los planes de recuperación de desastres) se definió la meta de generar instrumentos y entornos de prueba para tecnologías de AM y compartir los resultados que coadyuven a las instituciones a implementar arquitecturas de almacenamiento con enfoque de PD.

Es así como surgió la necesidad de diseñar un primer instrumento para documentar y, realizar un análisis del estado actual e identificar los requerimientos de almacenamiento masivo (AM) en instituciones donde se pretende realizar acciones de PD.

El instrumento es un cuestionario de 28 reactivos, para recabar información sobre el interés por el almacenamiento con enfoque de preservación, tamaño del acervo, tipos de contenido y formatos a preservar, crecimiento, infraestructura actual y requerida, disponibilidad, medios, respaldos y copias, esquemas de recuperación y problemas comunes en el almacenamiento.

El instrumento, difundido en poco más de ocho meses, fue respondido por instituciones de México y Latinoamérica. Una vez concentrada la información, se procedió a su análisis. El resultado fue la obtención de un panorama sobre el estado actual y los requerimientos del almacenamiento masivo.

¹ Grupo multidisciplinario e interinstitucional creado en el 2017 impulsado por la Coordinación de Innovación y Estrategia Digital de la Biblioteca Nacional (México) y el Instituto de Investigaciones Bibliográficas (IIB) de la Universidad Nacional Autónoma de México (UNAM) con el objetivo de promover la investigación, experimentación y formación de recursos en torno a la problemática de la Preservación Digital (PD). <https://preservaciondigital.iib.unam.mx/>

² Sistema de Información de Archivo Abierto (OAIS por sus siglas en inglés). Es un modelo que a través del CCSDS (Consultative Committee for Space Data Systems) se registra como norma a partir de 2003.

³ La National Digital Stewardship Alliance (NDSA) es un consorcio que trabaja en temas de preservación desde hace más de 12 años; respecto al componente de almacenamiento propone cuatro niveles para proteger, conocer, controlar y reparar los datos.

⁴ En el artículo “Cinco años del grupo de preservación digital. Reflexiones en torno a sus procesos” se detalla sobre las buenas prácticas de PD [6].

El interés que despertaron los resultados de este primer análisis, los convirtieron en el antecedente para conocer la situación que prevalece en las instituciones involucradas con respecto a las tecnologías de almacenamiento masivo para la preservación digital.

2. Desarrollo

Almacenamiento Masivo

Entiéndase el almacenamiento masivo como un conjunto de sistemas destinados a guardar grandes cantidades de datos, a corto, mediano o largo plazos, en soportes físicos como discos ópticos o duros, cintas magnéticas, unidades *Zip* y la nube, entre otros, con el objetivo de buscar, leer y recuperar información, garantizando su permanencia con base en políticas de preservación, así como estrategias y tecnologías. También se asocia con copias de seguridad y la recuperación de datos.

Dentro de los estándares y proyectos internacionales, se tiene el modelo *OAIS* y la alianza *NDSA-LDP*⁵, como grandes aportes para la *PD*. *OAIS* está dedicado a la gestión, la preservación y el almacenamiento de archivos. Es utilizado para garantizar que el contenido digital sea almacenado adecuadamente, permanezca completo y reproducible con el tiempo, por medio de actualizaciones periódicas de los medios y la migración de formatos, y cuente con mecanismos de salvaguarda, procedimientos de comprobación de errores y planes de recuperación ante desastres [1]. El *NDSA-LDP* es un grupo con más de 140 organizaciones, que busca establecer, mantener y anticipar la capacidad para preservar los recursos digitales nacionales (EE. UU.), como un beneficio para las generaciones presentes y futuras, creando niveles de *PD*, buenas prácticas y guías estructuradas en 36 actividades de preservación digital englobadas en 5 categorías y ponderadas en cuatro niveles con acciones específicas en *PD* [2].

Es importante señalar que la *PD* tiene requerimientos de almacenamiento masivo muy particulares, acordes con la estrategia de preservación, visión y misión, establecidas por cada institución.

Descripción del Instrumento (temáticas, muestra y recolección)

Las preguntas del instrumento surgieron conforme a las experiencias del Grupo de Almacenamiento Masivo y las instituciones representadas, además de la revisión

de artículos referentes al tema y los niveles de almacenamiento de la *NDSA-LDP*. Se utilizaron preguntas de opción múltiple, cerradas y abiertas, para conocer otras tecnologías o procesos no contemplados, logrando mejorar y robustecer el instrumento en cuanto a reactivos e instituciones participantes.

El instrumento contempló las secciones:

1. Interés del almacenamiento en materia de *PD*: tipo de institución, experiencias problemáticas y/o necesidades, importancia o desconocimiento del tema.
2. Acervo digital: tamaño, tipo de formatos y contenidos a preservar, crecimiento y prioridad.
3. Infraestructura: estado, espacio, medios de almacenamiento y respaldo, actual y futuro.
4. Personal: capacitado y calificado.
5. Procesos: documentación, formas de preservar, esquemas de respaldo, mantenimiento y recuperación, sincronización de información, método criptográfico, copias de seguridad y distribución geográfica.
6. Interoperabilidad: tipo y disponibilidad de información.
7. Esquemas de recuperación.

Participantes

El instrumento se aplicó al *GAM*, obteniendo una muestra de seis registros. Después, se distribuyó al Grupo de Preservación Digital (*GPD*) [3], creciendo a 12 registros, y, por último, se compartió con instituciones externas, con injerencia e interés en la *PD*. Las instituciones participantes están enmarcadas en sectores gubernamentales, asociaciones civiles, organizaciones educativas, así como entidades científicas, sociales, institucionales, tecnológicas y culturales, además de bibliotecas y centros de investigación, alcanzándose una muestra de 32 registros: 21 instituciones mexicanas (10 entidades de la UNAM) y 11 instituciones latinoamericanas (tres de Colombia, tres de Nicaragua, dos de Cuba, dos de Panamá y una de Costa Rica). Ver Anexo I.

Análisis en cuanto a acervo, materiales y formatos

En el análisis, un 75% de las instituciones participantes mostró interés en preservar por normatividad, mientras que el 25% restante investiga e implementa acciones por convicción. Con respecto al tipo de formatos y contenido, las instituciones prefieren preservar archivos históricos que de servicios, y preservar multimedios que archivos comprimidos. La mayor variedad de formatos y contenidos la reportaron las entidades participantes de la UNAM.

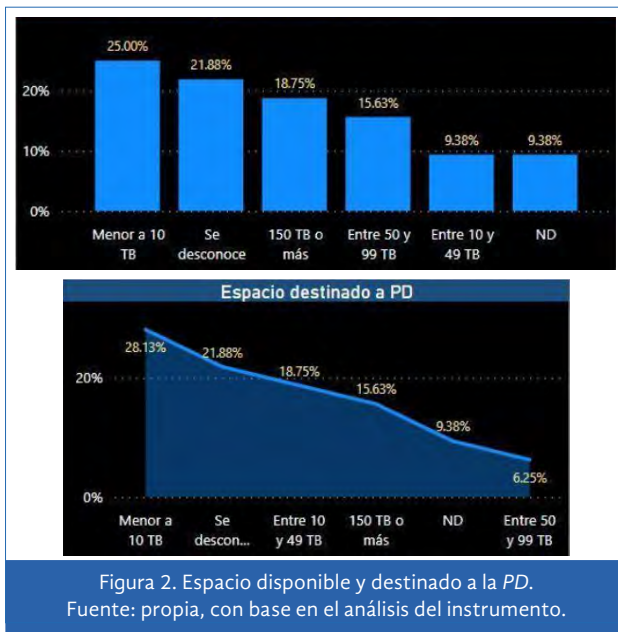
⁵ *Levels of Digital Preservation (LDP)* <https://ndsa.org/publications/levels-of-digital-preservation/>



Infraestructura y espacio de almacenamiento

El 37.5% de las instituciones participantes manifestó no contar con infraestructura para dedicarla a la PD, mientras que el 53% indicó que sí. De este segundo grupo, el 46.8% cuenta con equipos especializados, como servidores, unidades de almacenamiento, servicios en la nube, almacenamiento institucional remoto y sistemas de gestión de medios (MAM). De las entidades de la UNAM, el 72.7% cuenta con infraestructura dedicada y el 63.6%, especializada. El alto porcentaje obtenido por entidades de la UNAM puede deberse a la custodia de acervos importantes de carácter nacional, históricos y propios de la Universidad.

Sobre la infraestructura necesaria, fueron mencionados equipos para centros de datos o jerárquicos, ser-



vicios en la nube o distribuidos y sistemas completos de preservación, para el largo plazo y la alta disponibilidad, además de híbridos. Adicionalmente, se necesita personal debidamente capacitado para las labores de administración, captura, catalogación y digitalización.

En cuanto al espacio disponible y el designado para la preservación, se identificaron intervalos en ambos grupos de respuestas, desde los que cuentan con menos de 10TB, hasta los que pueden acceder a más de 150TB. Uno de los participantes indicó que su capacidad es ilimitada. Utiliza cintas LTO, que pueden incrementarse conforme sea necesario.

En 17 casos fue posible determinar la proporción del espacio que puede dedicarse a la preservación, en relación al disponible, encontrando que alrededor de una quinta parte puede dedicar la totalidad de su capacidad de almacenamiento y otra quinta, la mitad o más.

Tamaños de acervo y de archivos, disponibilidad y tasa de crecimiento anual

Los tamaños de acervo varían desde los mayores a 200TB (14.2%), hasta aquellos menores a 1TB (7.1%). La mayoría de los casos se encuentra en el rango entre 1 y 19TB (28.5%), aunque hay que considerar que el 32.1% lo desconoce o no contestó.

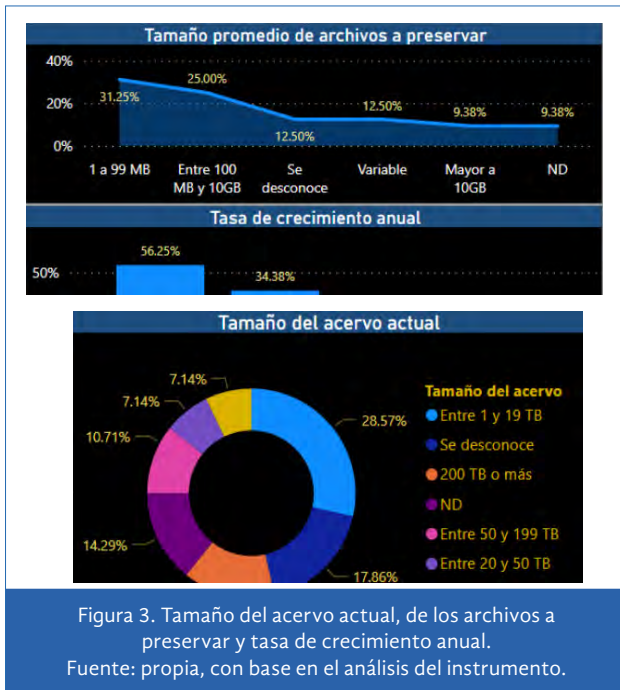
Al cuestionar el tamaño promedio de los archivos a preservar, se identificó que el 31.2% se encuentra en el rango de 1 a 99MB; el 25% entre 100MB y 10GB, y el 9.3% es mayor a 10GB.

En relación a la disponibilidad, el 55.1% manifestó que requiere que sea del 99.9%, mientras que el 37.9% lo estimó por arriba del 50%. Esto demuestra, nuevamente, que existe la necesidad de contar con almacenamiento de alta disponibilidad, así como para la preservación en el largo plazo.

En cuanto a las necesidades futuras, el 56.2% estima que su acervo crecerá anualmente en menos del 20%, mientras que el 34.3% considera que lo hará entre el 21% y el 40%. Sólo un 3.1% consideró que se duplicará.

Ubicación de la infraestructura de preservación e interoperabilidad

Si bien se recomienda que la infraestructura de preservación esté distribuida geográficamente, al analizar los resultados se encontró que el 48.3% sólo cuenta con almacenamiento en servidores propios y el 25.8%, con un esquema combinado, que incluye además una nube comercial o espacio en el centro de datos gestionado



ante la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) de la UNAM. En un solo caso se cuenta con un esquema de tres ubicaciones, que incluye servidores propios, el espacio en la DGTIC y un esquema híbrido (ver figura 4).

Acotando las respuestas a entidades de la UNAM, observamos que cinco de ellas ubican su almacenamiento únicamente en servidores propios y dos sólo en el centro de datos de la DGTIC, mientras que cuatro lo hacen en ambos. Cabe mencionar que, a pesar de la conveniencia de contar con almacenamiento en nubes públicas, por normatividad algunas instituciones no pueden hacerlo.

Con respecto a la interoperabilidad, necesaria para la comunicación entre los subsistemas de preservación, el

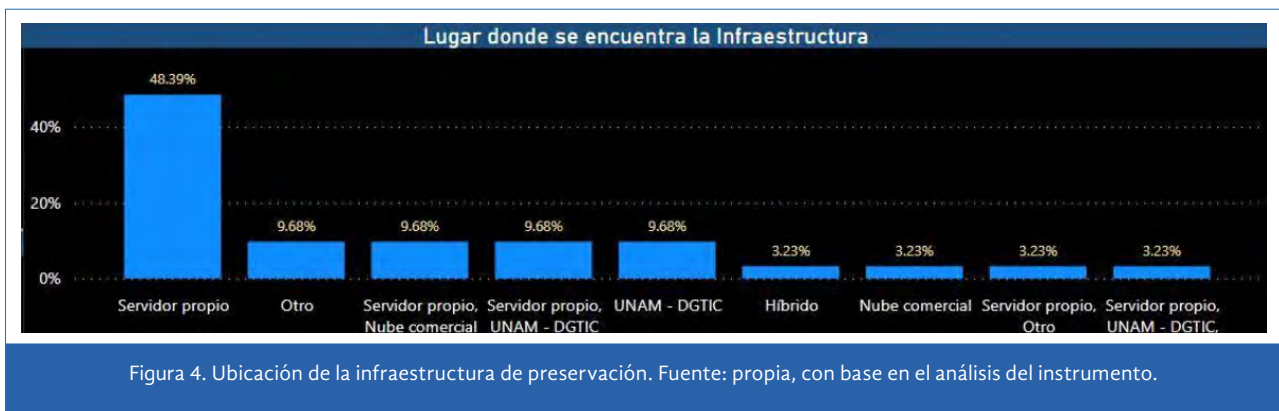
25% indicó que no es considerada, mientras que el 65.6% expresó que sí, a través de interfaces de programación de aplicaciones (API) y protocolos de almacenamiento.

Procedimientos en el almacenamiento

El análisis estableció que el 20% de las instituciones sólo tiene una copia de seguridad, casi la mitad (46.8%) cuenta con al menos dos copias y dos de cada 10 realizan tres copias. Por otra parte, casi el 10% asegura que no cuenta con copias de seguridad, lo que resulta en un riesgo muy alto para el acervo histórico y la continuidad de los servicios de esas instituciones. Según los niveles de PD, NDSA refiere que para alcanzar el nivel básico (proteger los datos) es necesario tener dos copias completas que no estén unidas, es decir, que se encuentren en dos dispositivos y con soportes heterogéneos [4]. En este sentido, los que realizan al menos un respaldo, lo hacen en la ubicación del contenido original, lo que se traduce en vulnerabilidad ante desastres naturales. Casi la mitad (46.6%) de los que realizan dos respaldos sí tienen una ubicación geográfica diferente para sus datos.

Resultó evidente la falta de normatividad al interior de las instituciones, para contar con un mecanismo escrito y formal que obligue a los responsables de los datos a realizar sus respaldos de manera procedimental. El 46% desconoce si los respaldos se hacen siguiendo un procedimiento, el 18% aseguró que no lo hace y sólo el 34% lo lleva a cabo de manera formal. Resulta importante decir que las entidades que realizan tres copias o más, aplican un procedimiento perfectamente establecido en los lineamientos de su institución.

La periodicidad de los respaldos debe contemplar el tipo de contenido, el uso, la importancia y el espacio disponible para tal fin. El análisis reveló que cuatro de cada 10 participantes desconoce la temporalidad con que se



realizan sus respaldos porque lo delegan al área de sistemas de su institución. Realizar respaldos de forma semanal y mensual fue lo más común, con un 17%. Sólo el 10% realiza copias diariamente. Un dato preocupante es que el 6.9% no realiza nunca una copia de seguridad.

Al analizar el medio de almacenamiento utilizado, se estableció a los *HDD*⁶, *CD*⁷, *NAS*⁸ y discos duros externos, como los dispositivos que más se utilizan para almacenar datos, como se observa en la figura 5. Resulta relevante que el almacenamiento en la nube sólo es utilizado por 10%. En gran medida se debe a que la mayoría son instancias educativas y de gobierno, que tienen restringida la salida de información de sus instalaciones por normatividad de protección de datos personales, sensibles y confidenciales (ver figura 5).

Otro aspecto a considerar es que sólo la tercera parte de las instituciones cuenta con esquemas de respaldo y recuperación. Otra tercera parte expresó que no lo tiene implementado y el resto lo desconoce. Esto indica que gran parte de estas entidades se encuentra indefensa para afrontar una pérdida de datos, que puede darse por daños en el hardware o el software e, incluso, por alguna intrusión que secuestre la información.

Problemas en el almacenamiento

Los problemas que más se presentan son la falta de mantenimiento a los dispositivos, escaso personal calificado y espacio insuficiente, que en conjunto abarcan el 57% de las contrariedades que se han presentado en las instituciones participantes. Estos datos son muy congruentes con la encuesta de la *NDSA*⁹ 2021, en que el 36% consideró que tiene personal insuficiente, el 17% subcontrata el servicio y el 7% no tiene personal técnico especializado [5].

3. Conclusiones

El almacenamiento masivo es un componente de la *PD* que requiere un análisis profundo para su implementación de forma adecuada. Este estudio es un primer esfuerzo por tener un panorama general de los factores que intervienen y que permitirán sensibilizar al personal involucrado de las instituciones para plantear

⁶ *Hard Disk Drive* (Disco Duro, por sus siglas en inglés).

⁷ *Compact Disc Recordable* (Disco Compacto Grabable, por sus siglas en inglés).

⁸ *Network Attached Storage* (Almacenamiento Conectado en Red, por sus siglas en inglés).

⁹ Results of the 2021 Staffing Survey, National Digital Stewardship Alliance. <https://osf.io/2rb7k>

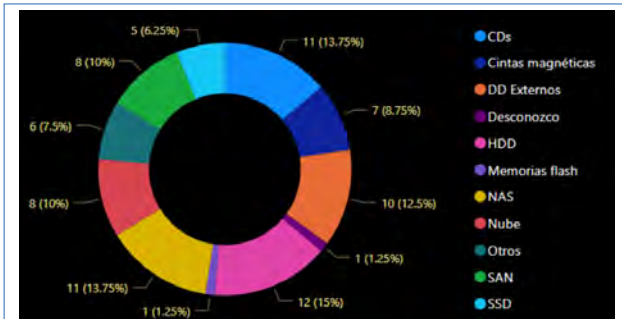


Figura 5. Medios de almacenamiento utilizados por las instituciones para almacenar sus datos.

Fuente: propia, con base en el análisis del instrumento.

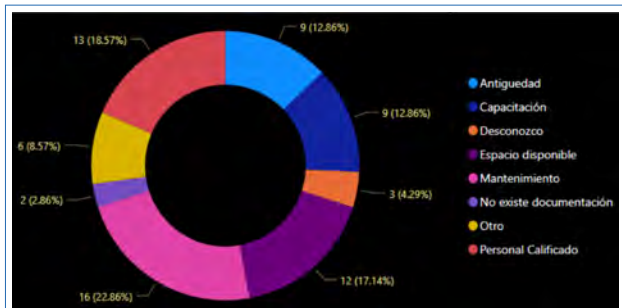


Figura 6. Problemas dectados en el almacenamiento de datos.

Fuente: propia, con base en el análisis del instrumento.

futuros proyectos, con una visión más completa de los procesos y los procedimientos de almacenamiento con enfoque de *PD*.

El instrumento de este estudio fue diseñado para aplicarse en las instituciones a través del personal que participa en temas de *PD*, sin embargo, sería muy importante adicionar preguntas que permitan profundizar en temáticas específicas del proceso de almacenamiento e integrar a más participantes. Asimismo, resulta necesario concientizar a los directivos de las instituciones sobre la importancia de invertir en personal e infraestructura adecuada y suficiente para garantizar el resguardo de la información, además de contar con un plan de almacenamiento a largo plazo, que contenga los procedimientos pormenorizados del almacenamiento de acervos digitales de la institución.

Finalmente, resulta conveniente conocer la infraestructura de almacenamiento con que cuentan las instituciones, así como el tamaño de los acervos y las estimaciones de crecimiento de sus archivos. Es un primer paso para entender el estado actual, las necesidades futuras y los retos que tendrán que superar para lograr la preservación efectiva de sus acervos, siendo el almacenamiento un aspecto crucial en cualquier estrategia de *PD*.

BIBLIOGRAFÍA

- [1] The consultative Committee for Space Data Systems, «Reference Model For An Open Archival Information System (OAIS),» Junio 2012. [En línea]. Available: <https://public.ccsds.org/pubs/650x0m2.pdf>. [Último acceso: 08 Febrero 2023].
- [2] D. A. Leija Román, «Universitat de Barcelona,» 2017. [En línea]. Available: https://www.tdx.cat/bitstream/handle/10803/454886/DALR_TESIS.pdf?sequence. [Último acceso: 2023 Febrero 22].
- [3] IIB-UNAM, «Grupo de Preservación Digital,» Instituto de Investigaciones Bibliográficas, UNAM, 2023. [En línea]. Available: <https://preservaciondigital.iib.unam.mx/txt/integrantes.html>. [Último acceso: 18 Febrero 2023].
- [4] NDSA, «Levels of Digital Preservation,» 2019. [En línea]. Available: <https://nds.org/publications/levels-of-digital-preservation/>. [Último acceso: 22 02 2023].
- [5] National Digital Stewardship Alliance (NDSA), «2021 Staffing Survey,» 07 Julio 2022. [En línea]. Available: <https://osf.io/2rb7k>. [Último acceso: 17 Marzo 2023].
- [6] I. Galina, R. Alvarado, B. Cabral, A. Castro, N. A. Manzanera Silva, A. López, A. Y. Ramírez y R. Sáenz, «CINCO AÑOS DEL GRUPO DE PRESERVACIÓN DIGITAL. REFLEXIONES EN TORNO A SUS PROCESOS,» 2022. [En línea]. Available: <https://boletinbm.iib.unam.mx/index.php/BBNM/article/view/363/639>. [Último acceso: 21 Marzo 2023].

Anexo 1

Institución	País
Archivo General de la Nación	Colombia
Caracol Televisión	Colombia
Instituto Popular de Cultura de Santiago de Cali	Colombia
Universidad Estatal a Distancia, Programa de Producción de Material Audiovisual	Costa Rica
Gabinete de Patrimonio Musical Esteban Salas	Cuba
Instituto Cubano de Radio y Televisión	Cuba
Archivo General de la Nación	México
Biblioteca Nacional de Antropología e Historia- Instituto Nacional de Antropología e Historia	México
Centro de Investigaciones sobre América del Norte	México
Dirección General de Cine	México
Equaltec (Empresa privada)	México
Fonoteca Nacional	México
Instituto Nacional de Musicología “Carlos Vega”	México
ONG	México
Universidad Autónoma de Yucatán, Fototeca Pedro Guerra	México
Biblioteca Nacional de México	México – UNAM
Centro Universitario de Teatro	México – UNAM
Dirección de Sistemas y Servicios Institucionales	México – UNAM
Dirección General de Actividades Cinematográficas	México – UNAM
Dirección General de Cómputo y de Tecnologías de Información y Comunicación	México – UNAM

Institución	País
Dirección General de Divulgación de la Ciencia	México – UNAM
Dirección General de Personal	México – UNAM
Dirección General de Repositorios Universitarios	México – UNAM
Instituto de Investigaciones Bibliográficas	México – UNAM
Instituto de Investigaciones Económicas	México – UNAM
Biblioteca Nacional de Nicaragua Rubén Darío	Nicaragua
Cinemateca Nacional	Nicaragua
Universidad Nacional Agraria	Nicaragua
Grupo Experimental de Cine Universitario	Panamá
Museo del Canal Interoceánico	Panamá

Fecha de recepción: 24 de marzo, 2023

Fecha de publicación: febrero, 2023

Cómo se cita

J. A. Contreras Benítez, M. G. Sánchez Mendoza, M. González González, O. Solís Garza, “Análisis del estado actual del almacenamiento masivo para la preservación digital”, *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero, 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

CRITERIOS PARA VALORAR SISTEMAS DE PRESERVACIÓN DIGITAL

DOI: 10.22201/dgtic.26832968e.2024.9.3

Ana Yuri Ramírez-Molina (yuri@unam.mx)
*Universidad Nacional Autónoma de México, Instituto de
Investigaciones Bibliográficas. Ciudad de México, México.*
ORCID: 0000-0001-8750-6976

Carlos Alberto Román Zamitz (carlos@fi-b.unam.mx)
*Universidad Nacional Autónoma de México, Dirección General
de Personal. Ciudad de México, México.*
ORCID: 0000-0002-4697-0040

Ivan Pérez Pineda (ivan.perez@correo.buap.mx) *Benemérita
Universidad Autónoma de Puebla, Biblioteca Histórica José
María Lafragua. Puebla, México.*
ORCID: 0000-0001-5201-1201

José Antonio Contreras Benítez (antoniocb7@gmail.com)
*Universidad Nacional Autónoma de México, Dirección General
de Personal, Dirección de Sistemas. Ciudad de México, México.*
ORCID: 0000-0001-8442-6030

José Antonio Salazar Carmona (tony@unam.mx)
*Universidad Nacional Autónoma de México, Instituto de
Investigaciones Bibliográficas. Ciudad de México, México.*
ORCID: 0000-0003-2032-4574

Juan Manuel Castillejos Reyes (visibilidad.web.dcv@unam.mx)
*Universidad Nacional Autónoma de México, Dirección General de
Cómputo y de Tecnologías de Información y Comunicación.
Ciudad de México, México.*
ORCID: 0000-0002-6714-0791

Maribel González González (pdigital@unam.mx)
Profesional independiente. Ciudad de México, México.
ORCID: 0000-0002-1430-7412

Teresa del Rocío González Melchor (melchor@unam.mx)
*Universidad Nacional Autónoma de México, Instituto de
Investigaciones Estéticas. Ciudad de México, México.*
ORCID: 0000-0002-9768-3447

Fecha de recepción: abril de 2023 • Fecha de publicación: febrero de 2024

www.ties.unam.mx

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

CRITERIOS PARA VALORAR SISTEMAS DE PRESERVACIÓN DIGITAL

Resumen

El presente artículo enumera una serie de criterios que ayudan a verificar o valorar el alcance de un Sistema de Preservación Digital (SPD), de acuerdo con las funcionalidades establecidas dentro del Modelo OAIS, además de las consideraciones señaladas en la propuesta para Auditoría y Certificación de Repositorios Confiables (TRAC) y el Modelo de Requerimientos para la gestión de registros electrónicos (Moreq), todos éstos agrupados con base en atributos de calidad de software.

Palabras clave:

Preservación digital, evaluación, sistemas de preservación digital, calidad de software.

CRITERIA TO VALUE DIGITAL PRESERVATION SYSTEMS

Abstract

This work lists a series of criteria that help to verify the scope of a Digital Preservation System (SPD) according to the functionalities established in the OAIS Model and recommendations from Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC); and Model Requirements for the Management of Electronic Documents and Records (Moreq), all of these, categorized according to the software quality attributes.

Keywords:

Digital preservation, assessment, digital preservation systems, software quality.

CRITERIOS PARA VALORAR SISTEMAS DE PRESERVACIÓN DIGITAL

1. Introducción

El trabajo que orientó el desarrollo de los criterios que aquí se describen surge en 2016, cuando la Biblioteca Nacional de México inicia labores de preservación digital y, para esto, lanza una convocatoria abierta y reúne especialistas de diversas dependencias de la UNAM e instituciones externas para conformar un grupo interdisciplinario e interinstitucional conocido como Grupo de Preservación Digital o *GPD*.

Uno de los objetivos del *GPD* fue analizar herramientas de *software* de apoyo en la preservación a largo plazo de objetos digitales, por lo cual, comenzó una investigación teórica, revisión de casos de éxito, procesos descriptivos para la documentación de hallazgos y entrevistas para determinar los factores o criterios necesarios para valorar el alcance de un Sistema de Preservación Digital o *SPD*. Esto permitió identificar cuáles son los procesos básicos requeridos por una institución para llevar a cabo el resguardo de objetos digitales a largo plazo.

La investigación teórica incluyó procesos descriptivos y de recolección de datos de los hallazgos, para determinar categorías de análisis, con lo cual se buscó: a) Entender los conceptos alrededor de la preservación digital; b) Determinar los instrumentos de evaluación para establecer los alcances de la preservación digital, y c) Identificar herramientas y plataformas de *software* existentes en el mercado que apoyaban los procesos de preservación digital, considerando tanto al *software* libre como al comercial.

También se revisaron casos de éxito para identificar métodos de evaluación aplicados por comunidades o instituciones que llevan a cabo procesos de preservación digital.

La recolección de datos también incluyó entrevistas con grupos de proveedores seleccionados, a partir de los siguientes criterios: a) Los que ofertan productos comerciales de preservación digital, señalados en los mercados como importantes y prometedores, de acuerdo con la investigación documental realizada previamente, y b) Proveedores con la capacidad de distribuir su producto en México y brindar el soporte técnico necesario.

El proceso de análisis consistió en la comparación y la codificación de los datos obtenidos de la investigación documental, las entrevistas y los instrumentos de evaluación revisados, con lo cual se obtuvieron categorías y criterios que se clasificaron e identificaron basados en diferencias y semejanzas de manera iterativa, aplicando comparación constante, a partir de los atributos de *software* que se definen desde el punto de vista de la ingeniería de *software*.

2. La preservación digital

La definición de preservación digital, de acuerdo con la Coalición para la Preservación Digital o *DPC* por sus siglas en inglés, es “una serie de actividades de gestión necesarias para asegurar el acceso continuo a los materiales

digitales por el tiempo que sea necesario” [1] y se centra en combatir factores de deterioro que producen la obsolescencia, para los cuales es posible aplicar estrategias como: la migración, la normalización, la emulación y/o el encapsulamiento:

- La migración consiste en la transferencia de una versión de *hardware* y/o *software* a otra, con el fin de preservar el contenido intelectual del objeto. Este proceso, a pesar de ser útil, puede ocasionar pérdida de características y/o apariencia en el objeto.
- La normalización considera la aplicación de estándares para organizar y codificar la información contenida en el objeto digital. Al respecto, existen propuestas como la *Declaración de Formatos Recomendados* para la construcción y almacenamiento de objetos digitales, publicada por la Biblioteca del Congreso de los EUA [2], [7] y [16].
- La emulación y/o encapsulamiento es un procedimiento que combina el *software* y el *hardware* para reproducir el contexto original y las características esenciales del objeto. La idea principal es que un formato antiguo u obsoleto funcione en un ambiente informático nuevo [3].

3. El modelo OAIS

La norma ISO 14721:2003 u *Open Archival Information System*, también conocido como *OAIS*, es el marco conceptual más importante para entender la preservación digital, ya que “define los procesos necesarios para preservar y acceder a los objetos de información de forma efectiva y a largo plazo, y establece un lenguaje común que los describe” [4].

Este modelo enumera los procesos que intervienen en la preservación de un archivo digital, que incluye [5]:

- Un ambiente, compuesto por tres actores: proveedores, consumidores y el administrador del archivo, apoyados

por una infraestructura que funciona para atender a un plan de preservación.

- La funcionalidad de la infraestructura está compuesta por seis procesos: Ingesta, Almacenamiento, Gestión de datos, Administración, Planeación de acciones para la preservación a largo plazo de los objetos y el acceso a éstos.
- La definición del objeto digital, compuesto por el archivo o archivos de computadora y los metadatos que incluyen las transformaciones controladas que sufre el objeto en pro de su permanencia con base en un plan de preservación. Por esto, es necesario tomar en cuenta las siguientes consideraciones:
 - La normalización y/o control de formatos o tipos de archivo de computadora sustentada en estándares.
 - La verificación o comprobación del estado en que se recibe el objeto durante la ingesta.
 - La definición del tipo y cantidad de metadatos que acompañan al objeto para apoyar las decisiones de almacenamiento para preservación y consulta.

Además, el modelo contempla estrategias de preservación, como la migración de la información a soportes y formatos nuevos, para lo cual se acompaña de un conjunto diverso de metadatos de identificación y control.

4. Los metadatos que acompañan al objeto digital

El modelo *OAIS* considera un conjunto de metadatos asociados a cada objeto para tener el control de éste, su accesibilidad y su preservación. De ahí que contemple desde los elementos que describen al objeto, hasta las operaciones y los cambios que va sufriendo por acciones de preservación. Con base en esto, es posible clasificar a los metadatos en los siguientes tipos: descriptivos, técnicos, administrativos y preservación (Ver Tabla 1).

Tabla 1. Clasificación de metadatos para la preservación digital

Tipo de metadatos	Información acerca de	Ejemplos	Descripción
Descriptivos	Contenido del objeto.	Título, autor, tema y fecha de publicación.	Detalla las características de la obra. Generalmente se obtienen de un sistema de gestión bibliográfico.
Técnicos	Archivo electrónico.	Tipo de archivo, tamaño, fecha de creación y tipo de compresión.	Detalla a la manifestación de la obra, por lo que incluye características y estructura del ítem digital.

Tipo de metadatos	Información acerca de	Ejemplos	Descripción
Administrativos	Usos del contenido del objeto, archivos y datos administrativos.	Estado de propiedad intelectual, términos de la licencia y propietario de los derechos.	Permiten controlar el acceso al objeto, con base en los derechos otorgados por el autor. También establece relaciones del objeto con el resto del acervo digital.
Preservación	Preservación, interoperabilidad y administración de recursos.	Verificación y evento de preservación.	Contiene datos para la verificación de persistencia del objeto y el registro de las transformaciones que ha sufrido.

5. Herramientas para preservación digital y Sistemas de Preservación Digital

Para identificar cuáles eran los tipos de herramientas de *software* para preservación digital que existen en el mercado o en la comunidad, se partió de su alcance, para eso se consultó la lista *COPTR* o *Community Owned digital Preservation Tool Registry*, disponible en <https://digitalpowrr.niu.edu/>, que contempla desde herramientas que apoyan parte o totalmente una función específica, descrita en el modelo *OAIS*, como podrían ser la ingesta, la validación y la verificación, por mencionar algunas, hasta aquellas que pueden ejecutar más de una funcionalidad o que dicen ejecutar la totalidad de las funciones descritas en el mismo modelo.

Así, a partir de lo encontrado, las herramientas se clasificaron en dos categorías de análisis. Las primeras son consideradas como *herramientas de uso específico*, pues tienen sólo algunas de las funcionalidades definidas en el modelo *OAIS*. Las segundas buscan soportar la totalidad de las acciones para funcionar como una plataforma de trabajo y control de las funciones establecidas en el modelo *OAIS*, consideradas como *Sistemas de Preservación Digital* o *SPD*, por contemplar la ingesta, el almacenamiento de paquetes de información, la gestión de metadatos, la administración y el control del proceso, la planeación de acciones para la preservación a largo plazo de los objetos y el acceso a éstos.

Un *SPD*, por lo tanto, idealmente debería contemplar en la ingesta acciones de empaquetamiento, verificación de integridad y generación de metadatos, entre otros. Y para el almacenamiento, contar con repositorios para el

resguardo de paquetes de preservación y paquetes de distribución, además de las capacidades para monitorear constantemente la integridad de estos repositorios e incorporar las funcionalidades que le permitan el control de cada una de las actividades mencionadas.

En este punto, las entrevistas con proveedores, identificados en el mercado como especialistas en *software* de preservación digital, fungieron como un primer intento para la identificación de características de evaluación, por lo que participaron especialistas en preservación digital, tecnólogos y bibliotecólogos, quienes hicieron preguntas desde tres puntos de vista: la preservación digital, aspectos técnicos y tecnológicos, y las necesidades y recursos desde un punto de vista documental.

Cabe señalar que durante las entrevistas que se realizaron con los proveedores identificados en el mercado, éstos mencionaron al modelo *OAIS* como la base de sus *SPD*, sin embargo, desde el punto de vista tecnológico y documental, era difícil unificar las características que se debían explorar en común.

A partir del trabajo realizado, la revisión de artículos y las experiencias documentadas, se generó un listado de herramientas de preservación digital, en el cual se clasificaron como específicas (*H*) y Sistemas de Preservación Digital (*SPD*), aquellas identificadas por su mención, con mayor frecuencia, durante el proceso de investigación llevado a cabo. Dicho listado describe cada herramienta, con base en la información encontrada en los sitios web y/o documentación de las mismas herramientas (Ver Tabla 2):

Tabla 2. Herramientas (H) y sistemas de preservación digital (SPD) identificados en el mercado

Nombre	Licencia de software	Descripción	Tipo
Rosetta	Comercial	Contiene un repositorio para objetos y herramientas para el monitoreo de los objetos.	SPD
Libsafe	Comercial	Contiene repositorio para objetos, herramientas para el monitoreo de los objetos, visualizador para la consulta de éstos y hardware para almacenamiento.	SPD
PiqI	Comercial	Microfilm para llevar lo digital a formatos analógicos, ya que graba en un formato analógico la imagen de la codificación en binario de los archivos.	H
Archivematica	Código abierto	Contiene un repositorio para objetos y herramientas para el monitoreo de los objetos digitales.	SPD
RODA	Código abierto	Contiene un repositorio para objetos y herramientas para el monitoreo de los objetos.	SPD
DAITSS	Código abierto	Contiene un repositorio para objetos y herramientas para el monitoreo de los objetos.	SPD
JHOVE	Código abierto	Ambiente que permite identificar el formato, validar y clasificar objetos digitales.	H
Fido / PRONOM	Código abierto	A través de comandos identifica los formatos de archivo.	H
Jpylyzer	Código abierto	Valida imágenes y extrae propiedades de los archivos JP2. El formato de archivo JPEG 2000 (formato de archivo JP2) almacena datos específicos de la aplicación (metadatos) en asociación con un flujo de código JPEG 2000, como la información necesaria para mostrar la imagen. El formato de archivo JP2 encapsula las características JPEG 2000 junto con otras piezas de información.	H
xcorrSound	Código abierto	Compara archivos de video.	H
Droid	Código abierto	Identifica formatos de archivo.	H
DPSP	Código abierto	Es una colección de <i>software</i> para la preservación digital.	SPD
Preservica	Comercial	Herramienta para controlar el proceso de preservación digital.	SPD

6. Los criterios de evaluación de herramientas de preservación digital

Durante el desarrollo de esta investigación se localizaron cuatro experiencias relacionadas con la evaluación de herramientas para preservación digital.

La primera es una encuesta realizada por universidades portuguesas, para lo que ellos llamaron “*Software de código abierto para repositorios de preservación digital*”. Se concentra en la revisión de repositorios, cuyos criterios de comparación son: 1) Estrategias de preservación digital; 2) Autorización/autenticación; 3) Capacidad de

búsqueda; 4) Vistas previas; 5) Capacidad de información; 6) Soporte multilingüe, y 7) Dinamismo de la comunidad de desarrolladores [6]. En esta encuesta se consideran expectativas que no necesariamente forman parte de las características funcionales de la herramienta a evaluar, como es el caso de los incisos 1 y 7.

La segunda referencia de evaluación se titula Revisión de distintas implementaciones para preservación digital: hacia una propuesta metodológica para la preservación y la auditoría de confiabilidad de (RI) [8] que, aunque carece de información respecto a los criterios

utilizados para la elección de herramientas, respalda su deliberación en el prototipo propuesto por la Universidad de Michigan y esboza los conceptos y los procesos a contemplar para la implementación de una plataforma de preservación digital, señalando la relevancia del TRAC o *Trustworthy Repositories Audit and Certification*, para asegurar la confiabilidad del repositorio.

Un tercer caso de evaluación es “*Selecting an Electronic Records Repository Platform at the South Carolina Department of Archives and History*” [9], que explica los pasos ejecutados por el Departamento de Archivos e Historia de Carolina del Sur para proteger sus documentos digitales y aclara que un servicio de respaldo no es suficiente, aún y cuando éste se base en la nube. Además, también toma como marco conceptual al modelo OAIS y considera estándares de metadatos como PREMIS, METS, Dublin Core y TRAC. Finalmente, realiza instalaciones para identificar los alcances funcionales de DAITSS, DPSP, Preservica y Archivemática, y los repositorios Fedora, DSpace, HYDRA, Islandora y ATOM. Se resalta la importancia de: 1) Documentarse con todo

aquello relacionado con la preservación digital; 2) Conocer estándares; 3) Establecer requisitos funcionales; 4) Hacer pruebas de *software*, y 5) Hablar con la gente sobre las lecciones aprendidas. [9]

El cuarto y último instrumento de evaluación consultado es el publicado por *The Council of State Archivists, CoSA* por sus siglas en inglés, que es el Consejo de Archivistas del Estado, en los Estados Unidos de Norteamérica. Esta institución publicó en su sitio web una hoja de cálculo titulada *Comparing Digital Preservation Tools*, al parecer elaborada por el estado de Indiana, que contiene el trabajo de revisión de 5 Sistemas de Preservación Digital: DPSP, Archivemática, DAITSS, RODA y Planets Suite. [10]

En esencia, el documento muestra una tabla con 8 grandes rubros (Ver Tabla 3): *submission (envío)*, *ingest (ingesta)*, *storage (almacenamiento)*, *management (administración)*, *preservation/normalization (preservación/normalización)*, *dissemination (diseminación)*, *standards compliance & interoperability (normas de cumplimiento e interoperabilidad)* y *usability (usabilidad)*. [11]

Tabla 3. Hoja de cálculo elaborada por el Consejo de Archivistas de los Estados Unidos para la comparación de herramientas de preservación digital. Tomado de <https://www.statearchivists.org/viewdocument/comparing-digital-preservation-tool> (Council of State Archivists 2018).

REQUIREMENTS	DPSP (http://dpsp.sourceforge.net/)	ARCHIVEMATICA (https://www.archivematica.org/wiki/Main_Page)	DAITSS (http://daitss.fcla.edu/)	RODA (http://roda.di.umi.nho.pt/?locale=en#home)	Planets Suite (http://sourceforge.net/projects/planets-suite/)
Disclaimer: This comparison was done around 2013-2014 and has not been updated, while the tools may have had. It is a matter of personal opinion and experience and should not be taken as a true fact. It may also contain errors and your verification and tests might inform you better. It did not intend to disparage any vendor or product and is an honest statement of the fact as perceived then. Use with caution. It is perceived more as a requirement gathering tool.					
Legend :					
y OR Green = Good					
n or Red = Bad					
yn or yellow= equal good/bad - needs monitoring / improvement / support					
n/a = not applicable					
? Or empty = unknown					
SUBMISSION					
Individual file	y	y	y	y	?
Folder	y	y	y	y	?
Sets of Folder and Files	y	y	y	y	?
checksum verification	y	y	y	y	?
Virus Check before submission	n	y	y	y	?
Submission manifest with integrity report SHA1 & MD5	y	y	y	y	?

INGEST					
Virus-checking upon ingest	y	n	y	y	?
checksum verification	y	n	y	y	?
Ingest Manifest with Integrity report	y	n	y	y	?
STORAGE					
Compliant with IOT environment?	?	?	?	?	
Security	n/a	n/a	n/a	n/a	n/a
Ongoing checksum verification	y	yn	y	y	?
Ongoing format monitoring and migration	n	yn	n	y	?
Storage Access monitoring and report	y	yn	y	y	?
Ongoing virus check?	y	n	y	y	?
MANAGEMENT					
User role: Submission	y	y	y	y	y
User role: Ingest	y	y	y	y	y
User role: Storage	y	y	y	y	y
User role: Setting and Management	y	y	y	y	y
User role: Preservation	y	y	y	y	y
User role: Access	y	y	y	y	y
Data Management	yn	y	y	y	y
Ongoing checksum verification	y	yn	y	y	y
Reporting	y	y	y	y	y
PRESERVATION/NORMALIZATION					
Original file always kept?	yn	y	y	y	?
Word Processing	y	y	y	y	?
Plain Text	y	y	y	y	?
Spreadsheets	y	y	y	y	?
Audio Formats	y	y	y	y	?
Moving Image / Video	n	y	y	y	?
Raster Images	y	y	y	y	?
Vector Images	y	y	y	y	?
Raw Image Format	n	y	?	n	?
Data Sets / Databases	y	n	?	y	?
Presentations	y	y	?	n	?
Email	y	y	y	n	?
Webpages	y	n	?	n	?
Geospatial	n	n	?	n	?
Mixed / Special Formats / Generic / Social Media	n	n	?	n	?
Softwares / or operating System Emulator	n	n	y	n	?
Creation of preservation and access copies	y	y	y	y	?
Stamp/Watermark Dissemination copies	yn	n	n	n	?
Ongoing analysis of submitted formats for support	n	yn	y	n	?
Generation/creation of preservation metadata		y	y	y	?
Ongoing checksum verification	y	n	y	y	?
Binary normalization	y	n	y	n	?
Automated quality Assurance of formats migration/normalisation	y	y	?	y	?
Opportunity for manual quality assurance of normalisation	y	y	?	y	?
support backup and restore activities	n	n	?	?	?
Integrity check of normalisation and report	y	n	y	y	
Possibility to choose/change pre-set normalisation formats?	n	yn	n	y	y

DISSEMINATION					
Format Directly Accessible (without Special Viewer)	n	y	y	y	?
Online/Web Accessible content	n	y	y	y	?
Searchable Content	y	y	y	y	?
Searchable Metadata	?	y	y	y	?
Supports retention schedules	n	y	?	y	?
Produces Access Copies	y	y	y	y	?
Interoperable dissemination(XML/API)	n	y	y	?	?
Ongoing checksum verification	y	n	y	y	?
Stamp/Watermark Dissemination copies	yn	n		?	?
STANDARDS COMPLIANCE & INTEROPERABILITY					
iRODS (micro services)?	n	y	y	yn	?
ISO 14721 (OAIS)	y	y	y	y	y
ISO 16363 (TRAC/TDR)	n/a	n/a	?	y	?
ISO 16175	?	y	?	n	?
EAD METADATA	n	y	?	y	?
EAC-CPF METADATA	n	y	?	n	?
PREMIS METADATA	?	y	y	y	?
METS	?	?	y	y	?
MIX	?	?	?	y	?
AGLS	y	n	?	n	?
Metadata type for each type of record?	y	n	y	y	?
others? MODS, MARC, Dublin Core, XML?	y	y	y	y	?
Application Programming Interface (API) or XML standards?	n	y	y	?	?
USABILITY					
Linux platform	y	y	y	y	?
Windows platform	y	yn	y	y	y
MAC platform	y	yn	y	y	?
Web platform	n/a	y	y	?	?
Packaged for use inside a Virtual Machine?	n	y	?	?	n
Very little RAM and little Processor use	y	n	?	?	y
Platform Independent Application (Java?)	?	n	y	y	?
Reliability in performing the functions	0.9	0.75	?	?	?
Documentation	y	y	y	y	?
Community Support	y	y	n	n	?
Flexibility/Customizability	y	y	?	yn	?
License (Free?)	y	y	y	y	y
License (OpenSource?)	y	y	y	y	?
Ease of installation/start	y	n	n	y	y
Ease in upgrading?	n/a	n/a	n/a	n/a	?
Difficult to use	yn	n	yn	y	?
Easy	y	n	yn	yn	?
All in one single application	y	y	y	n	y
Integration of multiple applications	y	y	y	y	y
Modifiable source code	y	y	n	y	?

Los rubros de evaluación podrían estar relacionados con el modelo OAIS, sin embargo, no se conoce documentación que lo confirme, más allá de que éstos concuerdan con elementos señalados en el modelo. En total, los rubros suman un poco más de 90 criterios y pueden ser calificados con una escala de: bueno, malo, necesita monitoreo, no aplicable y desconocido.

En general, podríamos decir que el archivo Excel agrupa el conjunto de requisitos considerados como necesidades del CoSA pues así lo señalan los creadores en el documento.

Además, aclaran que cuando una herramienta no cumple con cierto requisito, esto no debe clasificarse como bueno o malo, pero sí considerar que simplemente no es útil para las necesidades que busca la institución. Por ejemplo, dentro de la usabilidad existe un criterio que dice “MAC platform” y podríamos suponer que evalúa si el sistema puede ejecutarse en plataformas MAC, lo cual, en efecto, no es negativo, pero es importante para ellos. De hecho, el rubro de usabilidad, por lo general, integra elementos que buscan satisfacer necesidades muy concretas del CoSA. Este último caso confirma que

no sólo son importantes los elementos de OAIIS, además hay otros factores a considerar como características requeridas por la institución.

7. MoReq y TRAC

Existen dos instrumentos que sirven como referencia para la evaluación dentro del ámbito de la preservación digital. El primer instrumento, *MoReq* o Modelo de Requisitos para la Gestión de Documentos Electrónicos de Archivo (del inglés *Model Requirements for the Management of Electronic Records*), describe y define los requisitos para una correcta gestión de documentos electrónicos y, en consecuencia, los requisitos funcionales de un Sistema de Gestión de Documentos Electrónicos de Archivo [12], el cual es un “sistema de información que captura, administra y proporciona acceso a los registros a través del tiempo” [13].

MoReq describe los elementos esenciales que un sistema de registros debe tener para garantizar que los registros se gestionen correctamente, se pueda acceder a ellos en todo momento, se conserven durante el tiempo que sean necesarios y se eliminen adecuadamente una vez que el período de retención obligatorio ha expirado.

El segundo instrumento de evaluación es *TRAC* o *Trustworthy Repositories Audit and Certification Criteria*. Define elementos para realizar auditorías a repositorios digitales de conservación a largo plazo, por lo cual intenta reunir las mejores prácticas sobre los requerimientos necesarios para que un repositorio digital sea confiable en cuanto a la conservación de documentos.

Fue creado por el *Research Library Group* y la *National Archives and Records Administration (RLG-NARA)* e influenciado por el *Auditing and Certification of Digital Archives Project* del *Center for Research Libraries (CRL)* y por el *Catalogue of Criteria for Trusted Digital Repositories* del grupo de trabajo de *Nestor, Network of Expertise in Long-Term Storage and Digital Resources*. Establece la definición de un repositorio digital de confianza, mediante listas de comprobación formadas por 84 indicadores y está dividido en cinco apartados: infraestructura organizacional, gestión de los objetos digitales, tecnologías, infraestructura técnica y seguridad [14].

De estos dos instrumentos es posible identificar como conceptos sustanciales y exigibles durante la evaluación de un *SPD*, los siguientes [15]:

- Contar con información redundante, ya sea como una copia idéntica o de alguna otra forma, para evitar la pérdida de objetos o información.

- El control en los niveles de acceso al acervo, ya que es imprescindible respetar los derechos de autor de cada objeto resguardado.
- La elaboración de respaldos continuos, para garantizar la permanencia y el control de cambios sobre los objetos.
- El cuidado en la integridad de los objetos y los metadatos, a pesar de las políticas de migración y transformación que se integren dentro del acervo.
- La identificación de cada objeto, es posible acceder y controlar con precisión todo lo que sucede sobre el mismo.
- La utilización de estándares para la definición de metadatos.

8. Atributos de calidad de software y la definición de criterios

Como ya se mencionó, una de las perspectivas empleada para la revisión de los *SPD* durante las entrevistas fue la tecnológica y, en ese sentido, se identificó que los modelos *OAIIS*, *Moreq* y *TRAC*, están relacionados con los llamados requisitos funcionales dentro de la *Ingeniería de Software*, ya que estos instrumentos definen acciones o funciones que debe ejecutar el Sistema de Preservación Digital.

Desde esta misma perspectiva, también existen otros aspectos importantes, llamados no funcionales, que es el término empleado para referirse a las características que debe cumplir una aplicación, que no se relacionan con su funcionalidad, como el tipo de soporte técnico que el proveedor otorga, los sistemas operativos en los que se puede instalar el *software*, las herramientas con las que interactúa y el costo del mantenimiento, entre otros, que, como se vio en los diferentes estudios de evaluación que se analizaron, también son importantes para las instituciones que adquieren o desarrollan un *SPD*.

Desde este punto de vista es posible ampliar el espectro de la revisión e incluir, para el análisis de estos sistemas, una visión en términos de *Ingeniería de Software* que contempla características conocidas como “atributos de calidad”, las cuales son usadas frecuentemente para la evaluación y la medición del *software* en el ámbito tecnológico [16]:

1. **Funcionalidad:** describe el conjunto de funciones, tareas y/u objetivos que debe cumplir el *software* de acuerdo con el uso específico que éste tiene.
2. **Fiabilidad:** capacidades que presenta el *software* para recuperarse de fallas, tolerar problemas y mantener el nivel de prestación de servicio especificado.

3. Usabilidad: características del *software* para facilitar al usuario la comprensión de su funcionalidad con un mínimo de esfuerzo.
4. Eficiencia: capacidad del *software* para responder a una petición de usuario con la velocidad apropiada, bajo las condiciones y los recursos adecuados.
5. Mantenimiento: características del *software* que influyen para que el usuario invierta el mínimo esfuerzo para mantenerlo o mejorarlo.
6. Portabilidad: capacidades del *software* para adaptarse a diferentes entornos, proporcionadas por el propio *software*.
7. Compatibilidad: grado en el que un producto puede intercambiar información con otros productos. Contempla la coexistencia o la convivencia con otro *software* y la interoperabilidad.
8. Seguridad: grado en que un *software* protege la información y el acceso a los datos. Involucra: confidencialidad, integridad, responsabilidad y autenticidad.

Al revisar el espectro de cobertura de los ocho atributos de calidad, definidos para cualquier tipo de *software* y comparar estos con los diversos ejercicios de evaluación encontrados durante la investigación, se encontró que muchos elementos empleados en las encuestas podían agruparse de acuerdo con estos atributos y su agrupación otorgaba mayores beneficios para la valoración de Sistemas de Preservación Digital, pues dan claridad en cuanto a qué se puede revisar.

Por todo esto, se construyó una guía estructurada a partir de nueve secciones y 73 criterios, en donde las secciones están basadas en los ocho atributos de calidad de *software* y una sección más, que considera la forma en que se distribuye y/o comercializa, ya que este elemento no estaba contenido en ninguno de los criterios anteriores, pero era necesario.

Las secciones contienen criterios formulados como preguntas o frases que fueron elaborados a partir de los aspectos señalados en las necesidades identificadas en *OAIS*, *MoReq*, *TRAC* y el análisis de los tipos de metadatos necesarios para preservación (ver Tabla 4).

La intención es que el usuario de la guía vaya respondiendo afirmativa o negativamente dichos criterios al revisar la plataforma que desea adquirir o desarrollar, de tal manera que como resultado obtenga el alcance aproximado de ésta con respecto a los procesos del modelo *OAIS*.

Para identificar cuántos de los 73 criterios establecidos en la guía se relacionan con la funcionalidad definida en el modelo *OAIS* (que contempla seis procesos básicos) y así reconocer en qué porcentaje la guía valora cada requerimiento establecido por dicho modelo, las secciones y criterios se agruparon y contabilizaron a partir de los seis procesos de *OAIS*, tomando en consideración que es posible que uno o más de los criterios esté relacionado con uno o más procesos del modelo (ver tabla 4). Con lo cual se obtuvo que, de los 73 criterios, 41 valoran la Ingesta (56% de la guía), 27 la Gestión de Datos (37% de la guía), 26 la Administración (36% de la guía), 24 la planeación (33% de la guía), 42 el Almacenamiento (58% de la guía) y 28 el Acceso (38% de la guía) (ver Figura 1).

Además, la distribución de criterios también contempla requerimientos no funcionales que fueron asociados con *OAIS* y la preservación digital y sustentados en la valoración del proveedor del *SPD*, considerando el acceso a servicios de soporte y una descripción en cuanto a la posición de la herramienta en el mercado respecto a rendimiento, diseño y robustez.

La guía está publicada en el sitio web del GPD <http://www.preservaciondigital.unam.mx/txt/base-productos.html>, descargable como un archivo *.pdf* de libre acceso.

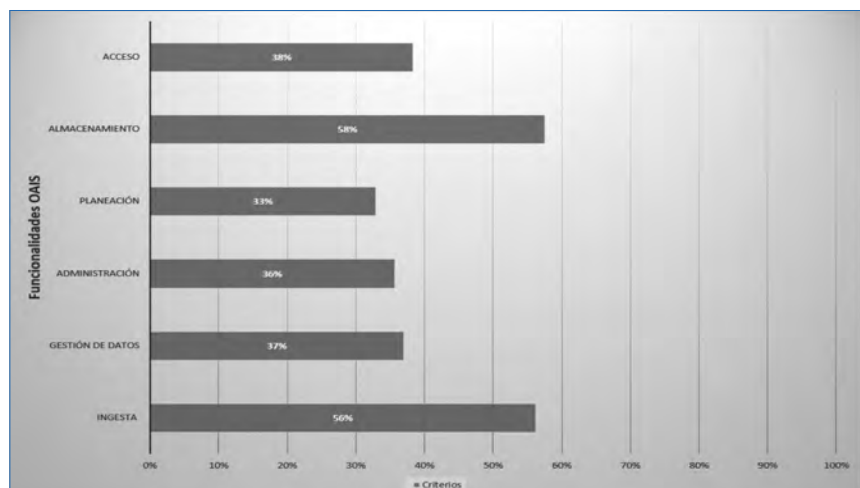


Figura 1. Porcentaje de distribución de los criterios en los procesos del modelo *OAIS*.

Tabla 4. Lista de distribución de los criterios en los procesos del modelo OAIS.

Ingesta		Gestión de datos		Administración		Planeación		Almacenamiento		Acceso	
#	Sección	#	Sección	#	Sección	#	Sección	#	Sección	#	Sección
1.3.1.1 al 1.3.1.3	Depósito y recepción	1.3.2.1 al 1.3.2.5	Gestión de datos	1.1.1 al 1.1.5	Administración	1.2.1 al 1.2.4	Planeación	1.3.3	Repositorio y/o almacenamiento	1.4.1 al 1.4.3	Acceso
1.3.2.1 al 1.3.2.5	Gestión de datos	1.2.1 al 1.2.4	Planeación	2.2 al 2.4	FIABILIDAD	2.1 al 2.4	FIABILIDAD	1.2.1 al 1.2.4	Planeación	2.1 y 2.4	FIABILIDAD
1.2.1 al 1.2.4	Planeación	2.1 al 2.2	FIABILIDAD	3	USABILIDAD	7	COMPATIBILIDAD	2.1 al 2.4	FIABILIDAD	3	USABILIDAD
2.1 al 2.2	FIABILIDAD	7	COMPATIBILIDAD	3.1.2 al 3.1.5	General	7.1.2 al 7.1.3	COEXISTENCIA (Paquetes OAIS)	4.1 al 4.4	EFICIENCIA	3.1.1 al 3.1.5	General
3	USABILIDAD	7.1.1 al 7.1.3	COEXISTENCIA (Paquetes OAIS)	3.2.1 al 3.2.4	Avisos y alertas	7.2.2 al 7.2.3	COEXISTENCIA (INGESTA/SIP)	5.1 al 5.5	MANTENIMIENTO	3.2.1 al 3.2.4	Avisos y alertas
3.1.3	General	7.2.1 al 7.2.3	COEXISTENCIA (INGESTA/SIP)	3.3.1 al 3.3.2	Ayuda	7.3.1	COEXISTENCIA (PRESERVACIÓN/AIP)	6.1 al 6.4	PORTABILIDAD	3.3.1 al 3.3.2	Ayuda
3.2.1 al 3.2.4	Avisos y alertas	7.3.1	COEXISTENCIA (PRESERVACIÓN/AIP)	6.1 al 6.3	PORTABILIDAD	7.4.1	INTEROPERABILIDAD	7	COMPATIBILIDAD	5.1 al 5.5	MANTENIMIENTO
3.3.1 al 3.3.2	Ayuda	7.4.1	INTEROPERABILIDAD	8	SEGURIDAD	8.1.1	CONFIDENCIALIDAD	7.1.2 al 7.1.3	COEXISTENCIA (Paquetes OAIS)	7	COMPATIBILIDAD
7	COMPATIBILIDAD	8	SEGURIDAD	8.1.1 al 8.1.2	CONFIDENCIALIDAD	8.1.1	CONFIDENCIALIDAD	7.2.2 al 7.2.3	COEXISTENCIA (INGESTA/SIP)	7.4	INTEROPERABILIDAD
7.1.1 al 7.1.3	COEXISTENCIA (Paquetes OAIS)	8.1.1	CONFIDENCIALIDAD	8.3.1 al 8.3.3	RESPONSABILIDAD	8.2.1 al 8.2.3	INTEGRIDAD	7.3	COEXISTENCIA (PRESERVACIÓN/AIP)	8	SEGURIDAD
7.2.1 al 7.2.3	COEXISTENCIA (INGESTA/SIP)	8.2.1 al 8.2.3	INTEGRIDAD	8.3.1 al 8.3.3	RESPONSABILIDAD	8.3.1 al 8.3.3	RESPONSABILIDAD	7.4	INTEROPERABILIDAD	8.1.1 y 8.1.2	CONFIDENCIALIDAD
7.3.1	COEXISTENCIA (PRESERVACIÓN/AIP)	8.3.1	RESPONSABILIDAD	8.4.1 al 8.4.2 y 8.4.4	AUTENTICIDAD	8.4.1, 8.4.2 y 8.4.5	AUTENTICIDAD	8	SEGURIDAD	8.3.1 al 8.3.3	RESPONSABILIDAD
7.4.1	INTEROPERABILIDAD	8.4.1 al 8.4.2 y 8.4.4	AUTENTICIDAD					8.1.1	CONFIDENCIALIDAD	9.3	DISTRIBUCIÓN Y/O COMERCIALIZACIÓN
8	SEGURIDAD							8.2.1 al 8.2.3	INTEGRIDAD		
8.1.1	CONFIDENCIALIDAD							8.3.1 al 8.3.3	RESPONSABILIDAD		
8.2.1 al 8.2.3	INTEGRIDAD							8.4.1 al 8.4.5	AUTENTICIDAD		
8.3.1 al 8.3.3	RESPONSABILIDAD							9.2 al 9.3	DISTRIBUCIÓN Y/O COMERCIALIZACIÓN		
8.4.1 al 8.4.5	AUTENTICIDAD										

9. Conclusiones

Uno de los principales retos con la aplicación de esta guía es llegar a identificar soluciones integrales para la preservación digital, donde se combinen y se aprovechen al máximo las cualidades de cada herramienta, con lo cual puedan cubrirse las diversas necesidades y requerimientos de las instituciones que pretenden instaurar soluciones tecnológicas para su estrategia de preservación.

La guía se hizo para que los interesados en la preservación digital identifiquen aspectos relevantes al intentar seleccionar un sistema como plataforma de gestión de sus procesos. Refleja los rubros esenciales del modelo OAIS y contempla las recomendaciones de TRAC y Moreq, así como los atributos de calidad necesarios en una herramienta de software.

Para utilizar la guía es necesario que la institución interesada cuente con un equipo humano con conocimientos vastos en tecnologías de la información y la comunicación; la gestión y la conservación de documentos y las ciencias de la información, como la bibliotecología, para comprender conceptos como: legibilidad, autenticidad, integridad y la importancia de los metadatos, además de entender cuáles son las necesidades y las capacidades de la organización, para así determinar adecuadamente qué valores son primordiales, dados los objetivos de preservación digital institucionales.

La guía no indica cuáles son los criterios indispensables para que un sistema de preservación digital se

considere útil en una institución. Tampoco determina la utilidad de un SPD y mucho menos lo califica como bueno o malo. Sin embargo, sí abre la puerta para el análisis del conjunto de elementos que habrá que tomar en cuenta, al tratar de seleccionar un sistema como éstos, basado en las necesidades de cada institución y los elementos básicos para hacer preservación digital.

Facilita las valoraciones de sistemas que hay en el mercado con los mismos criterios, donde cada institución dará mayor peso a los que considere más importantes, de acuerdo con sus necesidades. De igual forma, la guía está abierta a nuevas perspectivas, a partir de lo que cada usuario identifique.

Un paso importante y previo a la elección o elaboración de un SPD, es la definición de un plan de preservación, ya que en éste se detallan los recursos humanos, técnicos, financieros, legales y organizacionales con los que cuenta la institución para determinar la logística a seguir para establecer estrategias, programas, políticas, objetivos, acciones y prioridades, para llevar a la práctica la preservación digital.

Cabe aclarar que la guía de criterios básicos para valorar Sistemas de Preservación Digital no fue hecha para evaluar herramientas para la preservación digital de propósito específico, ya que éstas sólo se emplean en ciertos procesos o partes y no funcionan como plataformas para procesos de preservación.

Agradecimiento

Agradecemos a la Biblioteca Nacional de México por el impulso que ha dado a la preservación digital con la conformación del Grupo de Preservación Digital (GPD), lo cual permitió el trabajo colaborativo que se presenta en este artículo, y en el cual participaron especialistas de diversas entidades y dependencias de la UNAM así

como de otras instituciones encargadas del resguardo documental en México. También agradecemos el apoyo de la Mtra. Beatriz Zamora Ruíz de la Dirección General de Bibliotecas y Servicios Digitales de Información de la UNAM, por el proceso de revisión con respecto a las fuentes bibliográficas y la correcta citación.

BIBLIOGRAFÍA

- [1] "Digital Preservation Coalition." *dpconline.org*. Accessed: May, 5, 2019. [Online]. Available: <https://dpconline.org/handbook>
- [2] "Metadata Encoding & Transmission Standard." *loc.gov*. Accessed: February, 9, 2019. [Online]. Available: <http://www.loc.gov/standards/mets/>
- [3] Rivera Donoso, Miguel Angel, "Directrices para la creación de un programa de preservación digital," *Serie Bibliotecología y Gestión de Información*, no. 43, pp. 1-63, March, 2009. Fecha de consulta: 02 de septiembre de 2019. [Online]. Available: http://eprints.rclis.org/12989/1/Serie_N%C2%B043_Preservaci%C3%B3n_digital.pdf
- [4] Cruz Mundet, José Ramón and Díez Carrera, Carmen 2016, "Sistema de Información de Archivo Abierto (OAIS): luces y sombras de un modelo de referencia." *Investigación bibliotecológica*, vol. 30, no. 70, pp. 221-247. Accessed: Septiembre, 2, 2019. [Online]. Available: <https://dx.doi.org/10.1016/j.ibbai.2016.10.010>
- [5] "Preservation Metadata for Digital Objects: A Review of the State of the Art," *White Paper by the OCLC/RLG Working Group on Preservation Metadata*, June 31, 2001. Accessed: Oct, 18, 2017. [Online]. Available: http://www.oclc.org/resources/research/activities/pmwg/presmeta_wp.pdf
- [6] Rosa, Carlos André, Craveiro, Olga, and Domingues, Patricio, "Open Source Software for digital preservation repositories: A Survey." *International Journal of Computer Science & Engineering Survey (IJCSSES)*, vol. 8, no. 3, pp. 21-39, June, 2017. [Online]. Available: <https://arxiv.org/abs/1707.06336 DOI:10.5121/ijcses.2017.8302>
- [7] "Preservation Metadata Maintenance Activity." *loc.gov*. Accessed: February 9, 2019. [Online]. Available: <https://www.loc.gov/standards/premis/>
- [8] De Giusti, Marisa Raquel and Villareal Luján, Gonzalo, submitted for publication, "***Revisión de distintas implementaciones para preservación digital: hacia una propuesta metodológica para la preservación y la auditoría de confiabilidad de (RI) ***"
- [9] Brian, Thomas. "Selecting an Electronic Records Repository." *Journal of Contemporary Archival Studies*, vol. 2. pp. 1-21. Accessed: June 25, 2019. [Online]. Available: <http://elischolar.library.yale.edu/jcas/vol2/iss1/2>
- [10] "About Council of State Archivists." *statearchivists.org*. Accessed: June 30, 2019. [Online]. Available: <https://www.statearchivists.org/about/who-we-are>
- [11] "Comparing Digital Preservation Tools." *statearchivists.org*. Accessed: June 30, 2019. [Online]. Available: <https://www.statearchivists.org/viewdocument/comparing-digital-preservation-tool>
- [12] Enciclopedia cubana. 2011. "MoReq." *Ecured.cu*. Accessed: Marzo 25, 2018. [Online]. Available: <https://www.ecured.cu/MoReq#Propuesta>

- [13] "ISO 15489-1:2001: Information and documentation - Records management - Part 1: General." *iso.org*. Accessed: April 26, 2017. [Online]. Available: <https://www.iso.org/standard/31908.html>
- [14] Alcaraz, Rubén. (2013, Enero 3). Herramientas de auditoría para la preservación digital [Online]. Available: <http://www.rubenalcaraz.es/pinakes/preservacion-digital/herramientas-de-auditoria-para-la-preservacion-digital/>
- [15] Criterios básicos para valorar sistemas de preservación digital. 2020. [Online]. Available: <https://www.iib.unam.mx/files/iib/libros-electronicos/Criterios-Basicos-Sistemas-Preservacion-Digital.pdf>
- [16] "ISO/IEC/IEEE. 2010." *Sebokwiki.org*. Accessed: June 6, 2019. [Online.] Available: https://www.sebokwiki.org/wiki/ISO/IEC/IEEE_24765
- [17] "Recommended Formats Statement" *loc.gov*. Accessed: February 9, 2019. [Online.] Available: <https://www.loc.gov/preservation/resources/rfs/TOC.html>
- [18] Reference Model for an Open Archival Information System (OAIS): recommended Practice. Consultative Committee for Space Data Systems. 2012. [Online]. Available: <https://public.ccsds.org/pubs/650x0m2.pdf>
- [19] "Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects," A Report by The OCLC/RLG Working Group on Preservation Metadata, Jun, 2002. Accessed: October 18, 2017. [Online]. Available: https://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf
- [20] Lavoie, Brian, "The Open Archival Information System (OAIS) reference model: introductory guide" 2nd ed. *DPC Technology Watch Report* 14-02 October, 2014. [Online]. Available: <https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file>
- [21] Romero, Hermenegildo. (2009). Calidad de software e [PowerPoint slides]. Available: <https://es.slideshare.net/MeneRomero/calidad-de-software-11458973>
- [22] Trustworthy Repositories Audit & Certification: Criteria and Checklist. The Center for Research Libraries and Online Computer Library Center, Inc. 2007. [Online]. Available: http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf
- [23] Trustworthy Repositories Audit & Certification: Criteria and Checklist. *The Center for Research Libraries and Online Computer Library Center*, Inc. 2007. [Online]. Available: http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf
- [24] Yingxu Wang. *Software Engineering Foundations: A Software Science Perspective*, Auerbach Publications. [USA]: Taylor & Francis Group, 2008.

Fecha de recepción: 28 de abril, 2023

Fecha de publicación: febrero de 2024

Cómo se cita

A. Yuri Ramírez-Molina, et. al. "Criterios para valorar Sistemas de Preservación Digital", *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero, 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

**PRESERVACIÓN DIGITAL Y DERECHOS DE AUTOR, UNA
VISIÓN ANALÓGICA EN UN MUNDO DIGITAL: PROBLEMAS DE
REGULACIÓN EN LA LEY DE DERECHOS DE AUTOR EN ESPAÑA**

DOI: 10.22201/dgtic.26832968e.2024.9.4

Miguel Ángel Corado Guerrero (corado42@gmail.com)
Grupo de Preservación Digital y UC3M
Madrid, España.

www.ties.unam.mx

Fecha de recepción: febrero de 2023 • Fecha de publicación: febrero de 2024

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

PRESERVACIÓN DIGITAL Y DERECHOS DE AUTOR, UNA VISIÓN ANALÓGICA EN UN MUNDO DIGITAL: PROBLEMAS DE REGULACIÓN EN LA LEY DE DERECHOS DE AUTOR EN ESPAÑA

Resumen

En el presente artículo se tratarán los principales conflictos y frenos que tiene la legislación española en Derechos de Autor respecto a la Preservación Digital y sus estrategias. El primer freno consiste en la escasa actualización legislativa, que implica que la velocidad de avance tecnológico es mucho mayor que la adaptación y producción normativa del poder legislativo. Además de contemplar una redacción legal generalista, con limitaciones y escasas interpretaciones. En segundo lugar, se encuentra la presión de las grandes empresas que quieren superponer su visión al legislador, esta coyuntura afecta al articulado (art. 37 del *TRLPI*) y por ende al uso de estrategias de preservación. En tercer lugar, trataremos la protección tecnológica (*DRM*), la cual establece usos permitidos en los objetos digitales, restringiendo así toda estrategia que parta desde la copia del objeto digital, además de no contemplar las limitaciones establecidas en el art. 37 de la *TRLPI*. Finalmente, se tratarán los contratos que son formas de acceder a grandes colecciones y fondos, no obstante, su uso implica una aceptación de determinadas limitaciones y usos que imposibilitan, a los entes, el preservar esos objetos digitales.

Palabras clave:

Preservación Digital, Derechos de Autor, regulación, problemáticas, España.

DIGITAL PRESERVATION AND COPYRIGHT, AN ANALOGICAL VISION IN A DIGITAL WORLD: PROBLEMS OF REGULATION IN THE SPANISH COPYRIGHT LAW

Abstract

*This article will deal with the main conflicts and brakes that the Spanish Copyright legislation has with respect to Digital Preservation and its strategies. The first brake consists of the scarce legislative updating, which implies that the speed of technological progress is much greater than the adaptation and normative production of the legislative power. In addition to contemplating a generalist legal wording, with limitations and few interpretations. Secondly, there is pressure from large companies that want to superimpose their vision on the legislator, this situation affects the articles (art. 37 of the *TRLPI*) and therefore the use of preservation strategies. Thirdly, we will deal with the technological protection (*DRM*), which establishes permitted uses of digital objects, thus restricting any strategy that starts from the copy of the digital object and does not contemplate the limitations established in art. 37 of the *TRLPI*. Finally, contracts will be discussed, which are ways of accessing large collections and funds, however, their use implies an acceptance of certain limitations and uses that make it impossible for the entities to preserve those digital objects.*

Keywords:

Digital Preservation, Copyright, regulation, problems, Spain.

PRESERVACIÓN DIGITAL Y DERECHOS DE AUTOR, UNA VISIÓN ANALÓGICA EN UN MUNDO DIGITAL: PROBLEMAS DE REGULACIÓN EN LA LEY DE DERECHOS DE AUTOR EN ESPAÑA

1. Introducción

El objetivo principal del presente artículo es contextualizar la actual situación que viven las instituciones de memoria y educativas, como podrían ser archivos, bibliotecas, museos o centros educativos a la hora de establecer actividades de Preservación Digital.

Cuando estas instituciones efectúan estrategias tecnológicas de Preservación Digital para conservar los objetos digitales a largo plazo, se encuentran con una barrera legal: la Ley de Derechos de Autor o Propiedad Intelectual. En otras palabras, los avances tecnológicos en materia de Preservación y Conservación se topan con un freno en su aplicación: una legislación que no las contempla ni ampara.

El Estado concreto que se analiza es España, con la Ley de Derechos de Autor (Texto Refundido de Propiedad Intelectual, *TRLPI*, en adelante) [1], con su artículo 37 para vislumbrar cuales son las principales problemáticas que sufren las estrategias tecnológicas de Preservación Digital a la hora de establecerse en una institución de memoria o educativa. A modo de enumeración serán: una legislación desactualizada, la presión de los *lobbies*, el uso de la protección tecnológica (*DRM*), y finalmente las licencias y contratos.

Esta investigación centrada en el Estado español puede generalizarse y contextualizarse en otros países

angloparlantes o latinoamericanos en busca de aprendizajes a la hora de renovar sus legislaciones, de tal forma que la Preservación Digital, y su importancia, sea considerada por parte de las Asambleas legislativas de dichos estados.

2. Desarrollo

2.1 Legislación desactualizada, una visión analógica para un mundo digital

El primer conflicto y elemento por tratar es la legislación y su escasa actualización y adaptación al entorno digital y tecnológico. Como bien indica la *IFLA* [2], el papel de las bibliotecas en la sociedad es el de un emplazamiento para la conservación y protección del conocimiento. Es por esta misma razón que tienen reconocidas exenciones específicas en las leyes de Propiedad Intelectual española en el ámbito impreso.

Sin embargo, nos encontramos con el paradigma tecnológico y digital donde el poder legislativo debe dar una nueva respuesta. El problema ya es evidente: la legislación no avanza a la misma velocidad que el devenir tecnológico. El articulado español tiene un fuerte componente analógico que impide la integración de las nuevas potencialidades tecnológicas de la Preservación Digital. A esto se le puede añadir, señala Térmens [3], la

regulación simplista de la Preservación Digital que tiene el actual *TRLPI*. Este es uno de los nuevos retos que han de hacer frente bibliotecas y centros de información en la economía informatizada, expone Ramos [4].

Esta situación es expresada con claridad por parte de la *IFLA* [2], donde las legislaciones son redactadas de forma general, sin concretar mientras la tecnología se va desarrollando y las limitaciones son establecidas con muy poca posibilidad de interpretación, cosa que las hace muy difícilmente adaptables a la nueva tecnología que puedan implantarse en las estrategias de Preservación Digital. Asimismo, indican que esta constante legislativa hace que los articulados se vuelvan obsoletos al devenir tecnológico.

Bajo nuestro punto de vista, la legislación padece de unos puntos negativos que son mejorables, estos son: redacción generalista, limitaciones y no adaptación a la nueva coyuntura tecnológica. La legislación va atrasada al devenir tecnológico ya que para esto requiere tiempo, reflexión y prueba. Las soluciones legislativas al avance tecnológico son siempre una respuesta obsoleta.

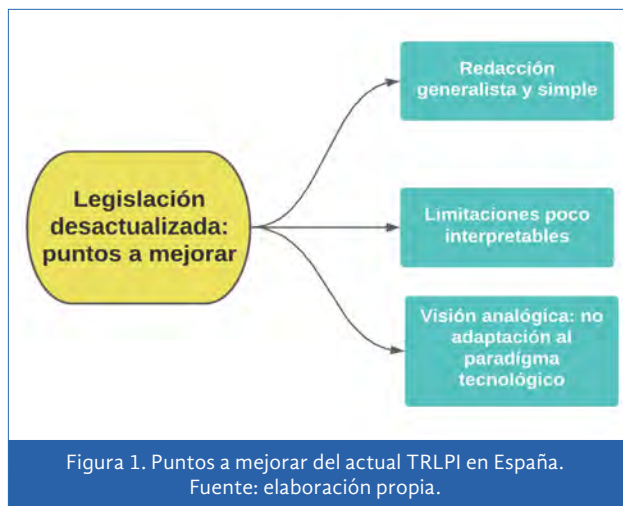


Figura 1. Puntos a mejorar del actual *TRLPI* en España.
Fuente: elaboración propia.

Esta opinión es compartida desde los inicios de esta disciplina, en un trabajo del *OCLC* [5, p. 38]: “*Research is needed to identify where current copyright protections inhibit digital preservation and how technical strategies might impinge on copyright laws*”. En cuanto al caso español, este se materializa en el artículo 37 del *TRLPI* [1], siendo su actualización escasa y no adaptada, las dos últimas actualizaciones (2007 y 2019) dan fe de esta situación. Se proponen, por lo tanto, revisiones periódicas que permitan contemplar los avances tecnológicos en materia de preservación y encargarlos así en el ordenamiento jurídico.

2.2 Multinacionales y grandes empresas, defensa contra la piratería y ataque a la Preservación Digital

El siguiente freno a la Preservación Digital es la posición de las grandes empresas que presionan al poder legislativo a la hora de legislar acerca de los Derechos de Autor, y de esta forma afecta también al art. 37 que contempla la Preservación Digital. En España, con el actual articulado, puede afirmarse que los intereses del conjunto de los ciudadanos por el mantenimiento y conservación de los objetos digitales para futuras generaciones quedan en un segundo plano al entrar los beneficios comerciales de las grandes empresas.

Este punto es sustentado por Térmens [3], que indica que España tiene una regulación simplista apoyada por las grandes corporaciones, bajo la bandera de la lucha contra la piratería. En el articulado pueden verse dichas presiones si se analiza el alcance de las limitaciones del art. 37 del *TRLPI* [1], en la protección tecnológica (*DRM*) y contratos mediante los cuales deja sin efecto real las limitaciones y excepciones a favor de archivos y bibliotecas aplicados en el *TRLPI*. Es más, puede señalarse que los Derechos de Autor son ejercidos por los titulares de los derechos, que, en la gran mayoría de los casos son propiedad de grandes editoriales o corporaciones [6].

En relación con lo anterior y en opinión de Voutsás [7], todas las legislaciones se establecen bajo la bandera de la lucha contra la piratería, pero su mala ejecución y poca consideración hacia la Conservación Digital ha desembocado en:

“una serie de pegotes, parches y enmiendas legales hechas a tontas y a locas, los cuales responden a impulsos del momento o a presiones e intereses de grupos, sin un entendimiento ni una visión integral y a largo plazo de lo que el problema representa” [7, p. 102].

Los titulares de los derechos tienen el legítimo derecho de defender sus intereses económicos. Pero no es menos importante que el objeto digital que defienden se preserve en el tiempo y de esta forma conservar y beneficiarse de la obra realizada por el autor. Esto último está por encima de una norma contra la piratería que no contemple el objeto digital como elemento legado a generaciones futuras. Por esto consideramos que los objetos digitales en España no tienen la regulación necesaria para su conservación a largo plazo.

Es más, este objeto digital suele estar bajo un contrato de las empresas con las instituciones de memoria a la hora de facilitar su contenido. Esta relación contractual impide su preservación o alteración, sumado a la protección tecnológica de *DRM* que paraliza toda estrategia de Preservación Digital. Dichos dos elementos se analizarán a continuación. Por mucha nueva estrategia de Preservación Digital que se quiera aplicar se topará inevitablemente con los intereses de empresas que, bajo el amparo de la lucha contra la piratería, limitan la aplicación de dichas estrategias.

2.3 Digital Right Management (*DRM*) y la protección tecnológica

El siguiente apartado profundizará en el *DRM* (*Digital Right Management*) es decir, la Gestión Digital de los Derechos. En otras palabras, es el apartado más tecnológico del presente artículo y a su vez una pieza clave por parte de la legislación a la hora de controlar los objetos digitales. Es pues, una de las problemáticas más importantes y que da sustento a los otros conflictos de regulación.

El término *DRM* es un concepto y dispositivo tecnológico a la vez, que permite un sistema de cifrado mediante la combinación de *hardware* y *software* encriptado, con ello se establecen usos permitidos por los titulares de las obras digitales [8]. Como añadidura, la legislación española instituye el *DRM* en el *TRLPI* en el artículo 196 del actual *TRLPI* y las limitaciones en el 197. En cualquier caso, la *DRM* limita a los usuarios y preservadores, indica Cordón y Arévalo [8], mediante las directrices que se muestran en la figura 2.

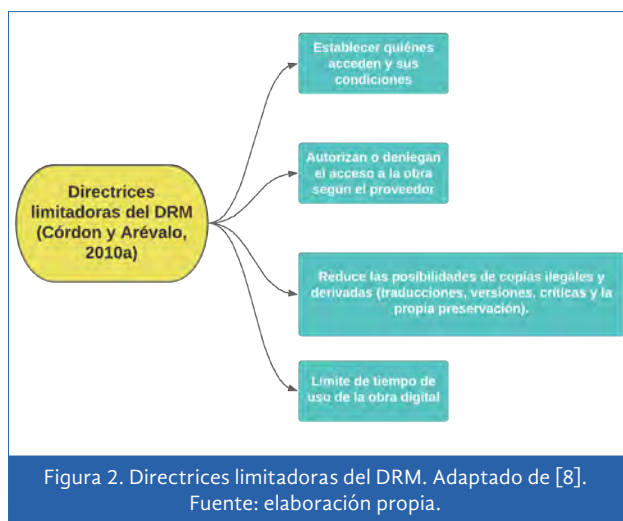


Figura 2. Directrices limitadoras del DRM. Adaptado de [8].
Fuente: elaboración propia.

Ramos [4] señala que la *DRM* es un derecho emergente que debe tratarse, pero su superposición está ahogando otros tantos derechos como el de lectura en una biblioteca o la intimidad en su acceso. En otras palabras, debido al *DRM* se restringen *de facto*, las limitaciones y excepciones que tienen los Derechos de Autor a favor de las bibliotecas y archivos e impiden el correcto ejercicio de los mismos emanados del art. 37 del *TRLPI* [1]. Esto en la práctica se traduce en una situación en la cual el titular de los derechos controla cómo utilizan los usuarios el contenido facilitado. Por ejemplo, una biblioteca presta un libro informatizado a un usuario, que en determinados días desaparecerá de su dispositivo y no podrá ser copiado o modificado.

El problema está en que toda estrategia de Preservación Digital, en primera instancia, pasaría por la copia. Esta duplicación está limitada o restringida y no se podría salvaguardar dicho elemento. Por ello, si un usuario quiere efectuar y hacer uso del art. 31 del *TRLPI* [1], para hacer uso del derecho de copia privada se verá en la tesitura de tener que utilizar su ingenio y conocimiento tecnológico para superar esta limitación técnica impuesta por el *DRM* [7]. Evidentemente, se está violentando la protección tecnológica, pero es esta misma la que impide el uso normal de las excepciones establecidas en la legislación. Es un conflicto de regulación que debe solventarse, debe mejorarse y clarificarse dicho uso. Se podría seguir el ejemplo de Finlandia que permite eludir las medidas tecnológicas según su propia Ley de Derechos de Autor [9].

2.4 Licencias y contratos, un escollo contra la preservación

El último elemento de conflicto dentro de la regulación son las licencias. Tratar las licencias o contratos es hablar de los sistemas que utilizan las bibliotecas para ofrecer diferentes colecciones a los usuarios y que puedan disfrutarlas a través de medios electrónicos. Pero detrás de este procedimiento, que facilita la lectura y permite un mayor acceso a los usuarios, hay un conjunto de prácticas que impiden la preservación del elemento digital a largo plazo.

Las potencialidades y problemáticas de este sistema son analizadas por Cordón y Arévalo [10], donde se establece que estos entes compran licencias para prestarlas a los usuarios, lo que funciona prácticamente como un libro físico: una licencia, un préstamo. Es evidente que

la editorial que facilita la licencia aplicará *DRM* al objeto licenciado para controlar en todo momento su uso y acceso. En cambio, esto está restringiendo y limitando toda estrategia de preservación futura.

Las editoriales distribuyen diferentes colecciones a las bibliotecas, como podría ser de un centro educativo. A su vez, estos entes buscan constantes actualizaciones de contenido, rapidez de acceso y funcionalidad para sus usuarios. Sin embargo, las editoriales y empresas ofrecen contenidos poco interesantes, y con varias limitaciones (mediante el *DRM*, de copia, impresión y descarga). Estas mismas restricciones impiden la libertad de uso al usuario o biblioteca, así como la aplicación de estrategias de conservación [10]. El equilibrio entre ambos actores es difícil de alcanzar, pero la realidad es que el sistema de licencias no facilita la salvaguarda del objeto digital.

La situación se complica si se sopesa el complejo trámite para alcanzar una licencia de suscripción, ya que no se adquiere realmente la obra. La editorial establece aún más limitaciones a la institución que lo adquiere, que debe firmar para acceder a dicha colección. Sin olvidarnos que las negociaciones para acceder a dichos contratos son extremadamente costosas, elemento que reduce los recursos posteriores destinados a esos mismos objetos digitales, véase la preservación señala Donoso, [11].

En otras palabras, se encuentra aquí un punto extremadamente conflictivo a la hora de efectuar la protección por parte de las bibliotecas. La biblioteca tiene una licencia con una editorial por unos determinados títulos, pero estos libros no serán de la institución sino de la propia editorial. Por esta razón impondrá las limitaciones que considere al objeto digital facilitado, entre ellas la copia. Aquí es interesante destacar el debate señalado por Voutssás [7] acerca de si es la editorial la que debe preservar al tener la propiedad sobre los mismos. Se puede ver tal situación en la figura 3.

Como se ha llegado a vislumbrar en este apartado, la protección contractual y las licencias que facilitan las editoriales van en contra de los principios de la Preservación Digital. Es cierto que la legislación actual del *TRLPI* permite exenciones para la conservación o copia. No obstante, esta se limita de forma repetida por los contratos entre bibliotecas y editoriales que impiden cualquier actividad de preservación.

3. Conclusión

Para finalizar, se debe destacar que esta investigación enfatiza los principales conflictos y frenos de regulación en la legislación española en materia de Derechos de Autor (*TRPLI*) a la hora de establecer la Preservación Digital en instituciones de memoria y educativas.

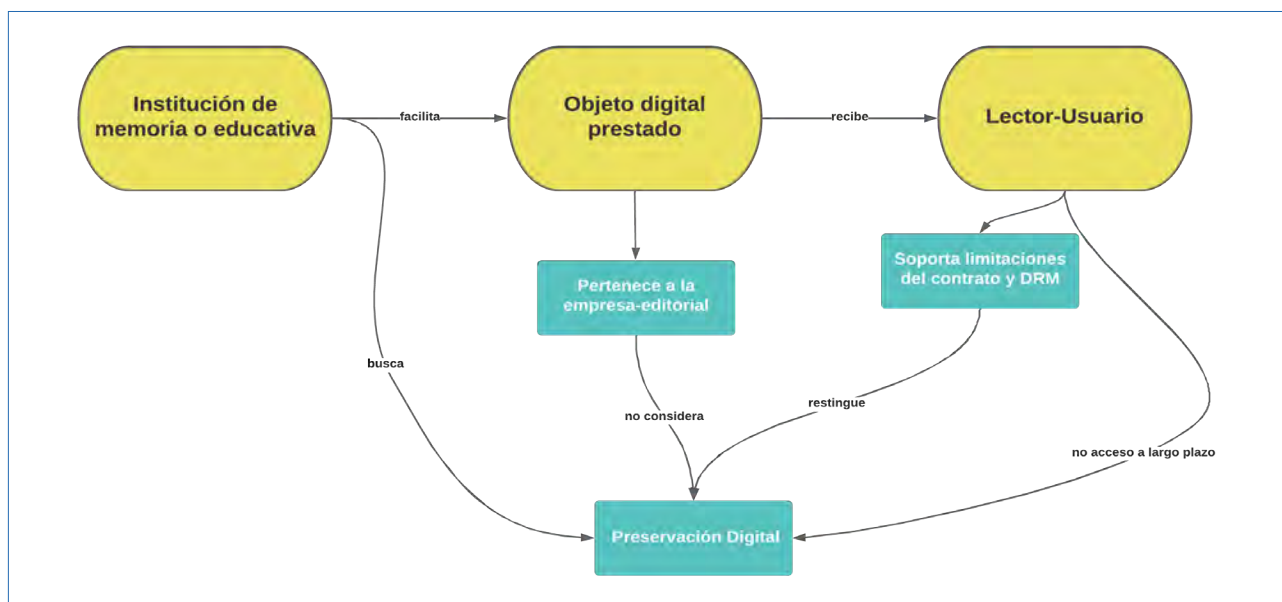


Figura 2. Relación entre la Institución, objeto digital y lector respecto a la Preservación Digital. Fuente: elaboración propia.

Todos los ítems tratados afectan en mayor o menor medida a las posibles actividades y estrategias que se quieran implantar. Es decir, causan una situación en la cual la institución tiene la tecnología para preservar el objeto digital a largo plazo pero su uso implicaría infringir la legislación actual.

Por todo lo anterior, puede concluirse qué elementos debe mejorar la actual ley y que pueden servir de ejemplo a otros ordenamientos jurídicos. Primero, lo más destacado es que la legislación en materia de Preservación Digital necesita actualizaciones más constantes para estar a la par con los avances en estrategias y herramientas de conservación. Asimismo, las empresas y corporaciones que abogan por legislaciones contrarias

a la preservación deben ser conscientes de que su no conservación es, a la postre, algo negativo para sus activos económicos y la integridad de sus objetos digitales a largo plazo.

Respecto a la protección tecnológica deberían contemplarse los entes de memoria y educativos como instituciones que pueden ser excepciones a dicha protección. Su función es preservar y conservar a largo plazo dichos objetos digitales para futuras generaciones, no para lucrar. Por último, los contratos y licencias deberían contemplar, en mayor medida, las estrategias de preservación hacia los propios objetos digitales, permitiendo una mayor interpretación del contrato suscrito y por ende la preservación.

BIBLIOGRAFÍA

- [1] España, Ministerio de Cultura. (1996, 23 de abril). Real Decreto Legislativo 1/1996, de 12 de abril, BOE núm. 97, Por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia. [En línea]. Disponible: <https://www.boe.es/buscar/doc.php?id=BOE-A-1996-8930>
- [2] IFLA, “Traducción al español del Tratado sobre Excepciones y Limitaciones para las Bibliotecas y archivos”, 2012. [En línea]. Disponible: https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/topics/exceptions-limitations/documents/traduccion_espanol_texto_propuesta_tratado_iflajuno2012.pdf
- [3] M. Térmens, “Gestión, uso y preservación de contenidos digitales: las necesidades técnicas frente a los derechos de propiedad intelectual”, en *V Congreso Nacional de Bibliotecas Públicas*, Gijón. Madrid: Ministerio de Cultura, 2010, pp. 37–44. [En línea]. Disponible: <http://hdl.handle.net/10421/5541>
- [4] F. Ramos, “La gestión de Derechos de Autor en entornos digitales, un reto para bibliotecas y centros públicos de información”, *Revista General de Información y Documentación*, vol. 12, núm. 1, pp. 257–277, 2002. [En línea]. Disponible: <https://core.ac.uk/reader/38822330>
- [5] Online Computer Library Center, “Trusted Digital Repositories: Attributes and Responsibilities”, *RLG-OCLC, Mountain View*, 2002. [En línea]. Disponible: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
- [6] J.C. Fernández-Molina, “Preservación Digital y Derechos de Autor: ¿un conflicto sin solución?”, en *V Congreso Nacional de Bibliotecas Públicas*, Gijón. Madrid: Ministerio de Cultura, 2010, pp. 43-50. [En línea]. Disponible: <http://hdl.handle.net/10421/5541>
- [7] J. Voutsás, “Factores tecnológicos, legales y documentales de la Preservación Digital”, *Investigación Bibliotecológica*, vol. 23, núm. 49, pp. 67–124, 2009. [En línea]. Disponible: <https://www.scielo.org.mx/pdf/ib/v23n49/v23n49a4.pdf>
- [8] J. Arévalo y J. A. Córdon, “El libro electrónico y los DRM”, *Anuario ThinkEPI*, pp. 249–253, 2010. [En línea]. Disponible: <https://thinkpi.profesionaldelainformacion.com/index.php/ThinkEPI/article/view/30513/16067>

[9] Instituto de Derechos de Autor, Mapa mundial de las leyes para la protección de la propiedad intelectual en el entorno digital. Madrid: Fundación SGAE, 2014.

[10] J. Arévalo y J. A. Córdon, “Las políticas de adquisición de libros electrónicos en bibliotecas: licencias, usos y Derechos de Autor” en *V Congreso Nacional de Bibliotecas Públicas*, Gijón. Madrid: Ministerio de Cultura, 2010, pp. 53-60. [En línea]. Disponible: <http://hdl.handle.net/10421/5541>

[11] M. Donoso, “Directrices para la creación de un programa de Preservación Digital.”, *Serie Bibliotecología y Gestión de Información*, núm. 43, pp. 1–63, 2009. [En línea]. Disponible: <http://eprints.rclis.org/12989/>

Fecha de recepción: 8 de febrero de 2023

Fecha de publicación: febrero de 2024

Cómo se cita

M. A. Corado Guerrero. “Preservación Digital y Derechos de Autor, una visión analógica en un mundo digital: problemas de regulación en la Ley de Derechos de Autor en España”, *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

LA HEMEROTECA NACIONAL DIGITAL DE MÉXICO: PERSPECTIVAS DE LA PRESERVACIÓN DIGITAL PARA EL PATRIMONIO DOCUMENTAL

DOI: 10.22201/dgtic.26832968e.2024.9.5

Alberto Castro Thompson (albertocastro@filos.unam.mx)
*Universidad Nacional Autónoma de México, Facultad de
Filosofía y Letras, Colegio de Bibliotecología y Archivología.
Ciudad de México, México.*
ORCID 0000-0001-7298-2742

Lisandro Pablo Olivares
(lspablo@comunidad.unam.mx)
*Universidad Nacional Autónoma de México, Biblioteca
Nacional de México. Ciudad de México, México.*
ORCID 0000-0002-6186-3745

www.ties.unam.mx

Fecha de recepción: abril de 2023 • Fecha de publicación: febrero de 2024

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

LA HEMEROTECA NACIONAL DIGITAL DE MÉXICO: PERSPECTIVAS DE LA PRESERVACIÓN DIGITAL PARA EL PATRIMONIO DOCUMENTAL

Resumen

La preservación digital es imprescindible en cualquier institución que valore sus recursos digitales. Desafortunadamente, aún existen múltiples repositorios, bibliotecas digitales y sistemas de información que, por distintas circunstancias, prescinden de una estrategia de preservación digital que contrarreste los efectos de la obsolescencia tecnológica, lo que supone un riesgo potencial para sus colecciones digitales.

Un elemento fundamental de la preservación es el proceso de creación de objetos digitales, pues en él se definen características relevantes de los archivos que pueden facilitar su administración a través del tiempo. En el caso de la digitalización, contempla la incorporación de procesos de selección de materiales y formatos de acuerdo con estándares y buenas prácticas, la inclusión de información de contexto suficiente, y la generación de contenidos confiables, auténticos y accesibles a largo plazo.

En el presente documento se describe el trabajo realizado por los departamentos de Preservación y Reprografía, Catalogación y Servicios de Información, de la Hemeroteca Nacional de México, y por el Departamento de Informática, adscrito a la Coordinación de Innovación y Estrategia Digital del Instituto de Investigaciones Bibliográficas de la UNAM, para la implementación de procesos de preservación digital en sus proyectos de digitalización y distribución de contenidos digitales, a través de la plataforma web Hemeroteca Nacional Digital de México.

Palabras clave:

TIC, preservación digital, Hemeroteca Digital.

THE NATIONAL DIGITAL NEWSPAPER ARCHIVE OF MEXICO: PERSPECTIVES OF DIGITAL PRESERVATION FOR DOCUMENTARY HERITAGE

Abstract

Digital preservation is essential in any institution that values its digital resources. However, there are multiple repositories, digital libraries and information systems that, due to different circumstances, do not have a digital preservation strategy that counteracts the effects of technological obsolescence, which represents a potential risk for their digital collections.

A fundamental part of preservation is the process of creating digital objects, since at this stage, relevant characteristics of the files that can facilitate their management over time, are defined. In the case of digitization, it involves the incorporation of appraisal and selection processes for materials and formats in accordance with standards and good practices, the inclusion of sufficient context information, and the generation of reliable, authentic, and long term accessible content.

This document describes the work carried out by the Preservation and Reprography, Cataloging, and Information Services departments, by the National Newspaper Library of México and by the IT department attached to the Coordination of Innovation and Digital Strategy of the National Library of México, for the implementation of digital preservation processes in its projects of digitization and distribution of digital content through the web platform Hemeroteca Nacional Digital de México.

Keywords:

TIC, digital preservation, Digital Newspaper Library

LA HEMEROTECA NACIONAL DIGITAL DE MÉXICO: PERSPECTIVAS DE LA PRESERVACIÓN DIGITAL PARA EL PATRIMONIO DOCUMENTAL

1. Introducción

La conformación de un plan o proyecto de preservación digital es una temática amplia y compleja. Incursionar en este tema incluye una larga lista de elementos que deben ser considerados y tomados en cuenta, con el fin de alcanzar un nivel óptimo para preservar lo digital con perspectivas hacia el futuro, como se observa en las definiciones de Miquel Térmens[1] y el proyecto *ICA e InterPARES* [2], así como en la carta de la *UNESCO* para la preservación del patrimonio digital [3].

En el caso específico de la Hemeroteca Nacional Digital de México (*HNDM*), la responsabilidad de preservar el patrimonio digital que resguarda tiene como origen el marco jurídico de la Ley General de Bibliotecas de México, expedida en junio de 2021[4]. En ella se especifica, entre otras cosas, instrumentar mecanismos de preservación del patrimonio documental editado y producido en México, tanto en formatos digitales como físicos, todo ello sin dejar de lado las actividades de organización, resguardo y difusión del referido patrimonio documental. Es decir, los recursos entregados en Depósito Legal deben estar disponibles para su consulta en todo momento, respetando los derechos de autor de cada una de las obras.

Es importante precisar que la preservación digital no se debe entender como sinónimo de resguardar todo bajo candados de seguridad, o bien, como instrumentos herméticos que encapsulan los recursos digitales, a

fin de inhabilitar su consulta y uso. Por el contrario, la preservación digital promueve la confiabilidad, la autenticidad, la integridad y la accesibilidad de los recursos digitales durante el mayor tiempo posible, situación que se convierte en todo un reto por encarar en un mundo tecnológico altamente cambiante y con un alto grado de obsolescencia, aspectos que son parte integral de la propia sociedad de la información.

2. Antecedentes de la *HNDM*

Desde su concepción, la Hemeroteca Nacional Digital de México (*HNDM*), ha sido un proyecto clave para la Hemeroteca Nacional de México (*HNM*) [5]. A lo largo de más de una década, en su desarrollo y evolución han participado diversos especialistas, quienes han realizado múltiples actividades de manera coordinada para poner a disposición del público en general las publicaciones periódicas y seriadas digitales que forman parte del patrimonio digital de México. Algo fundamental que recientemente se integró, tanto en la plataforma tecnológica como en la metodología de trabajo de la Hemeroteca Digital, es la inclusión de procesos de preservación digital, como parte de la cadena de producción, que comprenden el uso de estándares y la automatización de procesos.

En ese sentido, la proyección de la *HNDM*, como plataforma tecnológica, está más vigente que nunca,

gracias a las actualizaciones y las funcionalidades que se incorporan para facilitar el crecimiento y la consulta de las colecciones digitales, ello a pesar de la persistente amenaza de la obsolescencia tecnológica.

Es importante mencionar que, en la conceptualización de la *HNDM*, se identificaron cuatro principales factores a considerar sobre la transición de medios analógicos a digitales:

- I. La digitalización, como una forma de proteger los materiales analógicos, al limitar o evitar el acceso a los documentos impresos, disminuyendo a su vez el riesgo de deterioro o pérdida.
- II. Los documentos se distribuyen cada vez más en formatos digitales.
- III. Es esencial dar acceso a los materiales digitalizados, independientemente de la ubicación geográfica de los usuarios.
- IV. La tecnología de acceso a los recursos digitales es importante y requiere actualización continua.

Tomando en cuenta los factores previamente mencionados, la *HNDM* inició con dos objetivos principales:

1. Digitalizar los documentos impresos de mayor relevancia.
2. Poner los materiales digitalizados a disposición de usuarios y público en general, a través de una plataforma web.

Así, la *HNDM* se consolidó, desde su creación, como un repositorio virtual de publicaciones periódicas de carácter histórico que difunde el patrimonio hemerográfico del país (figura 1). Cuenta en la actualidad con 1,042 títulos, que representan más de nueve millones de imágenes de materiales digitalizados de diferentes épocas, comprendidas entre los años 1722 y 2010. Esto la convierte en el mayor repositorio de imágenes de periódicos y revistas en México.

A partir del 2019, se planteó una actualización importante para la *HNDM* que consistió en integrar procesos de preservación digital en los flujos de trabajo, de extremo a extremo, y la conformación del Sistema de Gestión de Recursos Digitales (*SGRD*). Ambos desarrollos fueron importantes para atender necesidades de administración y preservación de colecciones digitales.

Un aspecto clave para el desarrollo del proyecto es el fortalecimiento de la estructura tecnológica institucional, es decir, lo que el modelo de referencia *OAIS* (*Open Archival Information System*) denomina “servicios comunes” [6], los cuales se identifican como todas las aplicaciones (informáticas, de comunicación entre procesos, del sistema de información y de almacenamiento) así como los servicios proporcionados por los sistemas operativos, los servicios de red, la transferencia de archivos y la infraestructura relacionada con la seguridad, para proteger la información sensible ante riesgos y amenazas.

Es importante tomar en cuenta, adicionalmente, los criterios que recomienda el Comité Consultivo de Sistemas de Datos Espaciales (*CCSDS*, por sus siglas en inglés), llamados *TRAC* (*Trustworthy Repositories Audit and Certification*) [7], que están relacionados con la certificación de repositorios de confianza y consideran, principalmente, tres áreas: la infraestructura organizacional, la administración de los objetos digitales y la infraestructura tecnológica y de seguridad. En ese sentido, el desarrollo del proyecto *HNDM* busca integrar dichos criterios como parte fundamental para consolidar la preservación digital.

3. Digitalización, preservación y acceso

Es importante resaltar la continua confusión existente en la conversión de formatos analógicos a digitales, pues la digitalización, en muchas ocasiones, es considerada sinónimo de preservación digital. Si bien la digitalización en sí misma favorece el uso continuo de los materiales en el largo plazo [8], en realidad, para hablar de preservación digital, se necesita contemplar un conjunto de aspectos y

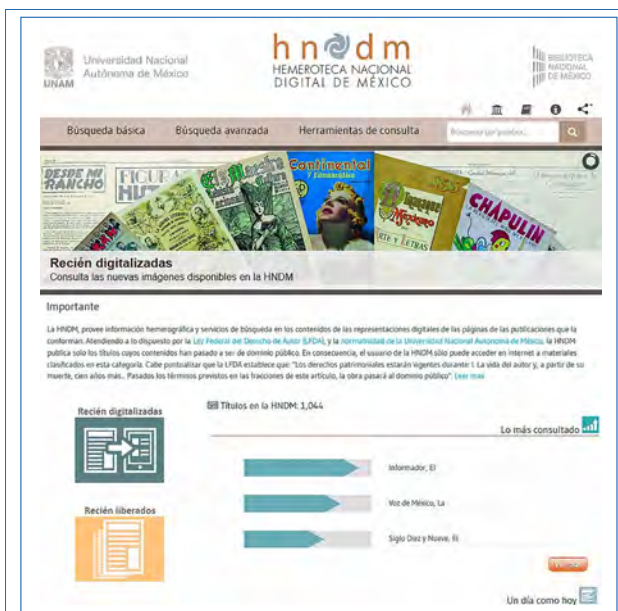


Figura 1. Hemeroteca Nacional Digital de México. Plataforma web.

técnicas favorables en los formatos digitales, que evite la obsolescencia y favorezca la integridad de los materiales, así como la accesibilidad continua en el tiempo a las colecciones digitales [9]. En otros términos, digitalizar es sólo el principio del ciclo continuo en el tiempo, pues cada objeto digital requiere de una vigilancia integral, constante y persistente.

Por lo anterior, es importante entender que la digitalización es el proceso de creación de archivos digitales mediante el escaneo o la conversión de materiales analógicos [10], que pueden ser libros, mapas y otros elementos impresos, aunque también se considera digitalización a la conversión de otros tipos de materiales distintos a los documentos, como imágenes, audios o videos a un formato digital.

En lo que respecta a los proyectos de digitalización, éstos se realizan normalmente en espacios de tiempo de corto o mediano plazo. Además, generalmente se facilita el acceso al material digitalizado a través de alguna plataforma informática. El principal objetivo de la digitalización es realizar una captura de calidad, bajo ciertos criterios técnicos que dependen del tipo de proyecto, en el entendido de que no es lo mismo digitalizar para preservar que digitalizar para distribuir.

Con respecto a la preservación digital, para la DPC (*Digital Preservation Coalition*) se define como “la serie de actividades administradas, necesarias para garantizar el acceso continuo a los materiales digitales durante

el tiempo que sea necesario. Son todas las acciones requeridas para mantener el acceso a los materiales digitales, más allá de los límites de la falla de los medios o cambios tecnológicos y organizacionales” [11]. Requiere políticas, planificación, asignación de recursos (fondos, tiempo y personal), tecnologías y acciones adecuadas para garantizar la accesibilidad, la representación precisa y la autenticidad de los objetos digitales [12].

La preservación digital incluye todos los tipos de recursos digitales, independientemente de su origen, tanto los nacidos digitales como aquellos digitalizados. Una buena práctica de la preservación consiste en tomar decisiones desde el momento en que se crean los recursos digitales. Dichas decisiones tendrán un impacto importante durante su ciclo de vida.

4. Incorporación de la preservación digital en la HNDM

Para la incorporación de los procesos de preservación en los flujos de trabajo de extremo a extremo, se definieron un conjunto de premisas a considerar, entre las que destacan:

- Digitalización ordenada y documentada.
- Catalogación de materiales digitalizados.
- Planificación e instrumentación de procesos de preservación digital.
- Inclusión de metadatos para una mejor preservación.
- Incorporación de estándares y buenas prácticas de digitalización.

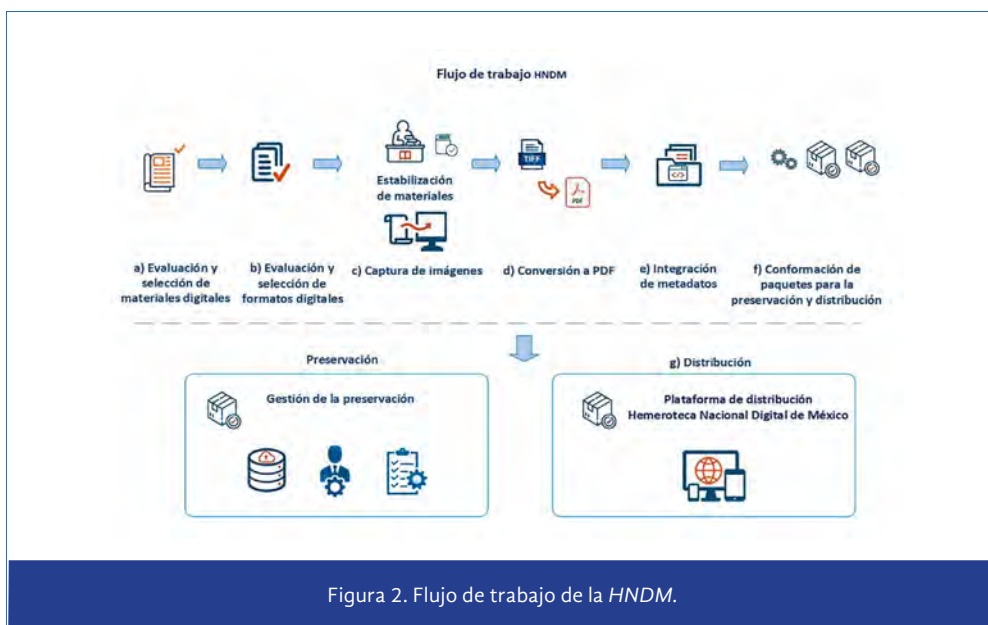


Figura 2. Flujo de trabajo de la HNDM.

- Automatización de procesos para la validación y la creación de paquetes de preservación y distribución.
- Optimización de flujos de trabajo.

A partir del establecimiento de estas premisas, se integraron procesos de preservación en el flujo de trabajo de la *HNDM*, como se muestra en la figura 2.

A continuación, se detalla el flujo de trabajo de la *HNDM*:

a) **Evaluación y selección de materiales.** Con respecto a la evaluación y la selección de materiales a digitalizar, la *HNM*, contempla los criterios y los factores siguientes:

- Alineación con la misión de la institución.
- Valor histórico, científico, cultural, administrativo y económico.
- La calidad, la unicidad y la representatividad.
- Autenticidad, integridad y origen.
- La demanda de la publicación por parte de los usuarios.

A los criterios descritos se suman otros más especializados, como la viabilidad técnica, la interoperabilidad y la responsabilidad, es decir, que los materiales a digitalizar estén bajo resguardo de la *HNM*.

Además, el estado legal de los materiales se considera un criterio muy importante para la digitalización, pues en este proyecto únicamente se digitalizan publicaciones del dominio público, publicaciones oficiales de los gobiernos estatales y publicaciones del gobierno federal, así como las publicaciones de las que se tiene la autorización expresa de los titulares de los derechos.

b) **Evaluación y selección de formatos digitales.** Para la evaluación y la selección de formatos digitales, fue necesario valorar las características técnicas y conciliar los requerimientos tecnológicos disponibles en el momento. Es importante recordar que el objetivo principal de la digitalización es producir imágenes digitales lo más parecidas a los materiales impresos originales: fotografías, mapas, planos, etcétera. Por esa razón, la calidad de la captura de imágenes digitales varía según su utilidad. En el caso de la *HNDM*, las imágenes deben integrar características de preservación, es decir, de alta calidad en la captura.

A partir de estas imágenes de alta calidad, se generan los archivos de distribución que se ponen a disposición de los usuarios en la plataforma web de la *HNDM*, razón por la cual es necesario elegir un formato adecuado.

Con respecto a la selección del formato para la preservación, y con base en la clasificación desarrollada

por la Biblioteca del Congreso de Estados Unidos [13], se tomaron en cuenta factores de sustentabilidad, como: mecanismos de protección, adopción del formato, documentación, divulgación, transparencia y apertura.

Con base en dichos factores, se optó por el formato *TIFF* (*Tag Image File Format*) [14] [15], que presenta las siguientes características:

- Es ampliamente adoptado para la generación de fotografías de alta calidad y digitalizaciones de alta resolución. Está implementado en cámaras digitales, escáneres y *software* de creación de imágenes.
- Ofrece algoritmos de compresión sin pérdida, por lo que los archivos *TIFF* conservan detalles y la profundidad de color de las imágenes originales.
- Funciona bien como imagen de origen, ya que se puede obtener una imagen con la mejor calidad posible.
- A pesar de ser un formato propietario, tiene suficiente documentación y está disponible en Internet.
- Es compatible con muchas cámaras, escáneres, sistemas operativos y programas de edición de imágenes.

Como formato de distribución de archivos se seleccionó el *PDF* (*Portable Document Format*) [16], por las características que se detallan a continuación:

- Presentación de la información (texto, imágenes, gráficos, etc.) de la misma manera, independientemente del dispositivo de visualización.
 - Formato de archivo universalmente compatible, basado en el formato *PostScript* y altamente portable.
 - Aceptado globalmente como el estándar convencional para el intercambio de documentos e información.
 - Respaldado por un gran ecosistema de sistemas operativos, aplicaciones de *software* y dispositivos móviles, que aseguran un acceso universal a los archivos.
 - Formato autocontenido.
 - Archivos optimizados para diferentes propósitos.
 - Documentación publicada y abierta.
- c) **Estabilización y captura de imágenes.** Todo material impreso, seleccionado para la digitalización, pasa por un proceso de revisión de su estado físico. En caso necesario, se realiza una estabilización del material, para evitar su desgaste o daño durante el proceso de captura. La estabilización se realiza de manera controlada y documentada. En esta etapa también se define la forma como será manipulado el material impreso y la forma correcta de hacer la captura, tomando en consideración sus dimensiones,

así como el tamaño del texto, las imágenes y otros contenidos relevantes [17].

Una vez verificado y estabilizado el material, se continúa con el proceso de captura de imágenes. Cada formato de archivo tiene propiedades específicas, que influyen en su apariencia, preservación y accesibilidad [18]. La captura es un proceso técnico especializado en el que se define la calidad de la digitalización. Por ello, las capturas de imagen se llevan a cabo de acuerdo con estándares y buenas prácticas, como los lineamientos *FADGI (Federal Agencies Digital Guidelines Initiative)* [19] sobre la captura de imágenes, tomando en cuenta los parámetros mínimos siguientes:

- Espacio de color.
- Profundidad de bit.
- Resolución.

En el caso de la *HNDM*, los parámetros seleccionados son: imágenes en formato *TIFF*, espacio de color *RGB*, con una profundidad de 8 bits por canal y una resolución de 300 píxeles por pulgada para textos de tamaño normal. Adicionalmente, las capturas realizadas pasan por un control de calidad en el Departamento de Reprografía de la *HNM*, a partir de los objetivos de captura establecidos previamente.

- d) **Conversión a PDF.** Una funcionalidad importante para la *HNDM* es la búsqueda, la localización y la consulta de textos dentro de los materiales. Para lograr dicha funcionalidad, a partir de la imagen *TIFF* se realiza una conversión a formato *PDF* y el reconocimiento óptico de caracteres (*OCR*, por sus siglas en inglés) para cada página generada en el proceso de captura. Como resultado, se obtienen los archivos con capacidad de búsqueda dentro del texto. Ambos archivos forman parte de la preservación a largo plazo.
- e) **Integración de metadatos.** Los metadatos son elementos fundamentales en cualquier proyecto de digitalización. Es una buena práctica recopilar tanta información, como sea necesaria, en el momento de la creación de los materiales digitales [20]. Por lo tanto, es recomendable la integración de metadatos sobre el objeto original y el digital creado, conteniendo, por ejemplo, parámetros de imagen digital, equipo utilizado, fecha y hora de creación, etcétera. En el caso de la *HNDM*, los dispositivos de captura de imágenes se configuran para agregar a los archivos el conjunto de metadatos técnicos de forma automática.

Otros metadatos que se agregan al archivo son los descriptivos. Éstos se obtienen del sistema de

información de la Hemeroteca y se incrustan tanto en las imágenes *TIFF* como en los archivos de distribución *PDF*. Todos los materiales a digitalizar deben estar correctamente catalogados, ya que el registro bibliográfico se agrega en un archivo tipo *MARC-xml*, conteniendo la descripción bibliográfica completa de la obra. Tanto la inserción de metadatos como la creación del archivo externo se realizan de forma automática con un programa de procesamiento por lotes. Un elemento importante en todo el proceso, es la asignación de identificadores únicos a los objetos digitales, los cuales sirven para localizar inequívocamente a los objetos, que en este proyecto se construyen a partir del número de sistema asignado por el sistema de información de la *HNM* y el código de barras asignado al material impreso.

- f) **Conformación de paquetes para la preservación y distribución.** Para la creación de paquetes fue necesario establecer una convención con respecto al nombramiento de archivos y la estructura de carpetas, la cual facilitó la automatización de los procesos de validación y verificación mediante el uso de un programa de cómputo. Este programa valida y verifica que las carpetas contengan los objetos digitales y los metadatos correspondientes, que el nombramiento de archivos sea el adecuado, que la estructura de carpetas sea válida y que los archivos *PDF* cuenten con capacidad de búsqueda en texto, entre otras validaciones.

Después de los procesos de validación y verificación se generan los paquetes de preservación y distribución. Los paquetes de preservación integran archivos de imágenes y materialidad en formato *TIFF*, archivos *xml* con la catalogación bibliográfica completa, archivos *PDF* de distribución y un archivo manifiesto con la suma de verificación de cada uno de los archivos. Una vez conformados, los paquetes pasan al sistema de almacenamiento de preservación, donde son resguardados y gestionados mediante procesos específicos para la preservación digital. Vale la pena mencionar que, para el resguardo de paquetes de preservación, se ha creado una estructura de almacenamiento estable con dos copias, y se continúa trabajando en la gestión de preservación, que comprende el desarrollo de actividades de seguridad, el monitoreo de almacenamiento, el control de acceso, la implementación de estrategias de preservación y la actualización de paquetes de distribución.

que precisaron de la capacitación del personal y la actualización de documentos guía que definen los objetivos de preservación digital en la institución. Adicionalmente, fue necesaria la integración de una infraestructura de almacenamiento confiable y el fortalecimiento de la infraestructura de red y de seguridad informática.

La preservación digital, como se mencionó, conlleva una serie de actividades planificadas que se deben definir e implementar, incluso antes de la creación de los objetos digitales, con el objetivo de tener archivos de calidad, auténticos, confiables y accesibles por el tiempo que sea necesario. En el caso de la *HNDM*, estas actividades incluyeron la adopción de estándares y buenas prácticas en los procesos de selección y evaluación de materiales, así como los formatos de los archivos. Con ello se generó una serie de procesos definidos y repetibles que ayudaron a realizar el trabajo de manera secuencial y eficiente.

En lo que se refiere a las actividades de captura de imágenes y proceso técnico especializado, además de establecer procedimientos con base en estándares, se aprovechó la vasta experiencia del personal en el uso de cámaras y escáneres para generar imágenes de alta calidad con fines y características de preservación.

En lo que respecta a los registros bibliográficos de las bibliotecas y centros de información, es claro que éstos son activos importantes y de gran valor. Por ello, se trabajó en la catalogación completa de los materiales a digitalizar y en la incorporación de dichos registros en cada uno de los paquetes de preservación generados. Otro

aspecto fundamental durante todo el ciclo de trabajo es la automatización en la creación de paquetes de preservación y distribución. Para ello, fue clave el desarrollo de programas de cómputo para minimizar posibles errores y, por otra parte, aplicar técnicas de validación y verificación de archivos que promovieron y potenciaron un incremento en la productividad de la *HNDM*.

Finalmente, es importante resaltar que el cumplimiento de los objetivos de preservación y distribución no es excluyente, es decir, se pueden generar paquetes de preservación de alta calidad con metadatos suficientes y confiables y, simultáneamente, paquetes de distribución con características adecuadas suficientes para la consulta del usuario final. En este caso específico, los paquetes a resguardar continúan su ciclo a través de la gestión de la preservación. Por otra parte, los paquetes de distribución son administrados por el sistema de gestión de la *HNDM* para, posteriormente, realizar su publicación y consulta desde el sistema.

Con los desarrollos instrumentados en el proyecto de la *HNDM*, se establece un precedente importante en lo referente a la administración de los recursos patrimoniales que se resguardan en la institución, al incorporar tecnologías para la preservación digital en los procesos de creación de colecciones digitales, digitalización y distribución. Así, la Hemeroteca Nacional de México avanza en su compromiso por preservar el patrimonio documental de México y en establecer mecanismos que den evidencia de su preservación.

BIBLIOGRAFÍA

- [1] M. Térmens, *Preservación Digital*, UOC, 2013.
- [2]. International Research on Permanent Authentic Records in Electronic Systems (InterPARES), *Desarrollo de políticas y procedimientos para la preservación digital*, vol. Temas fundamentales de preservación digital, México: Archivo General de la Nación, 2017.
- [3]. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, «UNESDOC Biblioteca Digital,» 2003. [En línea]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000229034_spa [Último acceso: 24 de marzo, 2023].
- [4]. Cámara de Diputados del H. Congreso de la Unión, *Ley General de Bibliotecas*, Ciudad de México, 2021.
- [5]. Hemeroteca Nacional de México, «*Hemeroteca Nacional Digital de México*,» [En línea]. [Último acceso: 23 de marzo, 2023].
- [6]. International Organization for Standardization, «ISO 14721:2012 Space data and information transfer systems — Open archival information system (OAIS) — Reference model,» [En línea]. [Último acceso: 10 de enero, 2023].
- [7]. Consultative Committee for Space Data Systems (CCSDS), «Trustworthy Repositories, Audit & Certification: Criteria and Checklist,» 2007.

- [8]. M. M. Mena Múgica y A. González Crespo, «Una imagen, mil palabras. La digitalización como estrategia de preservación digital,» de *Archivos electrónicos: textos y contextos II*, Puebla, México, Benemérita Universidad Autónoma de Puebla, 2013, pp. 15-24.
- [9]. P. Conway, «Handbook for Digital Projects: A Management Tool for Preservation and Access,» *Northeast Document Conservation Center*, 2000.
- [10]. Digital Preservation Coalition, «Digital Preservation Handbook,» [En línea]. Available: <https://www.dpconline.org/handbook/glossary> [Último acceso: 10 de enero, 2023].
- [11]. Digital Preservation Coalition, «Digital Preservation Coalition,» [En línea]. Available: <https://www.dpconline.org/digipres/what-is-digipres>. [Último acceso: 10 de enero, 2023].
- [12]. Oxford Bodleian, «Libguides, Digital Preservation, What is digital preservation?,» [En línea]. Available: <https://ufs.libguides.com/c.php?g=1113411&p=8118647> [Último acceso: 10 de enero, 2023].
- [13]. Library of Congress, «Sustainability of Digital Formats: Planning for Library of Congress Collections,» [En línea]. Available: Sustainability of Digital Formats: Planning for Library of Congress Collections. [Último acceso: 23 de marzo, 2023].
- [14]. International Organization for Standardization, «ISO 12639:2004 Prepress digital data exchange — Tag image file format for image technology (TIFF/IT),» 2017.
- [15]. International Organization for Standardization, «ISO 12234-2:2001 Electronic still-picture imaging — Removable memory — Part 2: TIFF/EP image data format,» [En línea]. Available: <https://www.iso.org/standard/29377.html> [Último acceso: 23 de marzo, 2023].
- [16]. «PDF Association,» 31 de enero, 2023. [En línea]. Available: <https://www.pdfa.org/resource/iso-32000-pdf/>
- [17]. M. Casey, «Media Preservation and Digitization Principles,» 17 de marzo, 2022. [En línea]. Available: https://scholarworks.iu.edu/dspace/bitstream/handle/2022/27446/casey_principles_final_20220326_.pdf [Último acceso: octubre, 2023].
- [18]. Digital Preservation Coalition, «Preserving Raster Images, DPC Technology Watch,» [En línea]. Available: Preserving Raster Images.
- [19]. Federal Agencies Digital Guidelines Initiative, «FADGI Guidelines,» [En línea]. Available: <https://www.digitization-guidelines.gov/>
- [20]. T. Traczyk, W. Ogryczak, P. Pałka y T. Sliwinski, *Digital Preservation: Putting it to work*, vol. Studies in Computational Intelligence, Springer.

Fecha de recepción: 1 de abril de 2023

Fecha de publicación: febrero de 2024

Cómo se cita

A. Castro Thompson, L. Pablo Olivares. “La Hemeroteca Nacional Digital de México: perspectivas de la preservación digital para el patrimonio documental”, *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

PRESERVACIÓN Y ACCESO A OBJETOS DIGITALES FUERA DE LÍNEA: EXPERIENCIAS EN LA FILMOTECA DE LA UNAM

DOI: 10.22201/dgtic.26832968e.2024.9.6

Gerardo León Lastra (gleonl@unam.mx)
*Universidad Nacional Autónoma de México,
Filmoteca. Ciudad de México, México.*

Manuel Comi Xolot (mcomi@hotmail.com)
*Universidad Nacional Autónoma de México,
Filmoteca. Ciudad de México, México.*

Gustavo Lucio José (lfgustavo@unam.mx)
*Universidad Nacional Autónoma de México,
Filmoteca. Ciudad de México, México.*

Luis Felipe Maciel Mercado (l_f_mm@hotmail.com)
*Universidad Nacional Autónoma de México,
Filmoteca. Ciudad de México, México.*

Fecha de recepción: abril de 2023 • Fecha de publicación: febrero de 2024

<https://www.ties.unam.mx/>

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

PRESERVACIÓN Y ACCESO A OBJETOS DIGITALES FUERA DE LÍNEA: EXPERIENCIAS EN LA FILMOTECA DE LA UNAM

Resumen¹

Se comparte la estrategia seguida en la Filmoteca de la UNAM para la preservación y acceso a largo plazo de ejemplares digitales de obras cinematográficas. La estrategia se diseñó durante la planeación de la integración del Laboratorio Cinematográfico de Restauración Digital (LCRD) de la Filmoteca. El plan incluyó el diseño y desarrollo del sistema de Control Logístico del Acervo Filmico (CLAF) para la gestión de activos digitales (MAM por sus siglas en inglés) en un almacenamiento en dos niveles —“en línea” y “fuera de línea”— que identifica y relaciona los ejemplares cinematográficos con sus contenedores, y estos con su ubicación espacial. Este diseño facilitará adaptar CLAF y sus procedimientos a un almacenamiento jerárquico para automatizar múltiples procesos de preservación y acceso.

La preservación del acervo digital de la Filmoteca peligró a corto plazo, dado que su dimensión ya alcanzada ha tornado imposible la ejecución manual, frecuente y periódica de la suma de comprobación de su fijeza. Para eliminar esta amenaza, se describe con detalle la metodología que facilitará la automatización de la comprobación periódica de la fijeza mediante la indispensable integración al LCRD de una librería de cintas.

Palabras clave:

Preservación digital, suma de comprobación de fijeza, almacenamiento jerárquico, librería de cintas, migración de datos.

¹ En este artículo se emplean las cursivas para términos en inglés, las **negritas** al definir o dar nuestra interpretación de un término y subrayaremos para enfatizar palabras.

PRESERVATION AND ACCESS OF OFF-LINE DIGITAL-OBJECTS: EXPERIENCES AT UNAM'S FILM ARCHIVE

Abstract

The strategy followed at UNAM's Film Archive for the long-term preservation and access of digital moving images is described. While planning the film digitization and restoration laboratory (LCRD for its Spanish acronym), this strategy included the design and development of a media asset management system (CLAF for its Spanish acronym) using 2 tier storage – “on line” and “off line”. CLAF identifies and establishes relationships amongst archive assets, its containers and their physical location. With this design, it will be relatively easy to adapt CLAF and its procedures to three levels of tiered-storage in order to automate processes of preservation and access.

Given the size already reached by Filmoteca's digital archive, its short-term preservation is under menace, since it is now impossible to manually execute with frequent periodicity its fixity check sum. To get rid of this threat, a tape library must be integrated into our LCRD lab. This integration will be relatively easy to do given the methodology followed and which is described in detail.

Keywords:

Digital preservation, fixity checksum, tiered storage, tape-library, data migration.

PRESERVACIÓN Y ACCESO A OBJETOS DIGITALES FUERA DE LÍNEA: EXPERIENCIAS EN LA FILMOTECA DE LA UNAM

1. Introducción

El último trimestre de 2014, la Filmoteca integró su “Laboratorio Cinematográfico de Restauración Digital” (*LCRD*) e inició la digitalización de obras cinematográficas de su acervo fílmico preservándolo en cintas magnéticas digitales. Desde el inicio del proyecto se identificó que después de algunos años de digitalizar películas se tornaría indispensable integrar una librería de cintas magnéticas para garantizar la preservación y el acceso a los productos de este laboratorio.

En enero de 2024 el acervo digital de la Filmoteca acumulaba 780 horas digitalizadas correspondientes a 1,601 títulos. Una sola copia, de las dos que se preservan, ocupaba más de 790 *terabytes* almacenados en 302 cintas *LTO-6*. Estas cifras seguirán creciendo hasta rebasar un estimado de 10 *petabytes* al concluir la digitalización del acervo fílmico —o más de 40 *petabytes* si aumentáramos la resolución de fotogramas de 2K a 4K—, como es conveniente.

Para gestionar el acervo fílmico de la Filmoteca y previo a la inminente integración del *LCRD*, utilizamos *software libre*, en la Filmoteca iniciamos en 2013 el diseño y el desarrollo en grandes fases o módulos del

sistema para el “Control Logístico del Acervo Fílmico” (*CLAF*). Este es un sistema híbrido que apoya la gestión de ejemplares en soporte fílmico y en soporte digital²; que utiliza códigos de barras (*CB*) para identificar los contenedores de ejemplares —latas, estuches de cintas, cintas, cajas, etc.— y utiliza otro *CB* para su ubicación a modo de coordenadas espaciales en las diferentes bóvedas de preservación. Además, *CLAF* permite visualizar en línea los *proxys* de las películas digitalizadas en baja resolución, el levantamiento del inventario físico y el control de proyectos de digitalización [1].

Desde su integración y hasta enero del 2024, el *LCRD* consta de una red local de 10Gbps que comparte almacenamiento de 73TB tipo *SAN* en *RAID-6* (120TB crudos) con dos escáneres de película cinematográfica, dos unidades de lectura/escritura (*i/o* por sus siglas del inglés) de cinta lineal abierta de sexta generación (*LTO-6* por las siglas del consorcio *Ultrium* [2]), estaciones de trabajo y

²El soporte digital son los bits que conforman cualquier objeto digital y debe distinguirse de su medio físico de almacenamiento que irremediamente es temporal [4, pp. 124-126]. Contrastando con el hecho de que el soporte digital puede preservarse por tiempo indefinido si se toman las acciones requeridas.

servidores virtualizados. A la fecha, CLAF gestiona más de 14,694 ejemplares³.

2. Datos y metadatos

Base de datos de CLAF

En [1] se expone un comparativo de modelos relacionales de las entidades principales o “nucleares”: el de CLAF, el modelo de referencia propuesto por la FIAF [3], y el empleado por el Instituto Británico del Film (BFI por sus siglas del inglés). Nuestro modelo identifica las obras cinematográficas y describe las características físicas y lógicas de los correspondientes ejemplares disponibles, incluyendo datos de su proveniencia y aspectos legales de propiedad. Todos los metadatos de contenidos cinematográficos y los datos para la gestión —como grupos de usuarios y privilegios, tareas de ubicación/recolección de contenedores, bitácoras e históricos de préstamos, etc.— se almacenan en una base de datos SQL, a excepción de los metadatos de fijez de ejemplares digitales cuyo manejo se describe más adelante.

Identificación de contenedores

En la Filmoteca empleamos el término **contenedor** para referirnos genéricamente a latas para guardar rollo(s) de filme y a cintas magnéticas LTO para almacenar objetos digitales. Utilizamos un entero de 32 bits para identificar cada contenedor de manera única. Habiendo planeado la futura integración de una librería de cintas, el formato del código de barras para identificar cintas y sus estuches, se hizo acorde a la especificación del consorcio LTO Ultrium [5]. Este código incluye la generación LTO a que pertenece la cinta [6] para facilitar su manejo empleando librerías de citas con unidades de i/o de diferentes generaciones de LTO.

Ubicación de contenedores

La Filmoteca cuenta con 16 bóvedas con estantería para preservar los ejemplares fílmicos y digitales. En CLAF denominamos **nicho** a un volumen claramente delimitado por postes y charolas o cualquier tipo de separador del estante (ver figura 1).



Figura 1. Tres pilas de film en 1 nicho.
Fuente: elaboración propia.

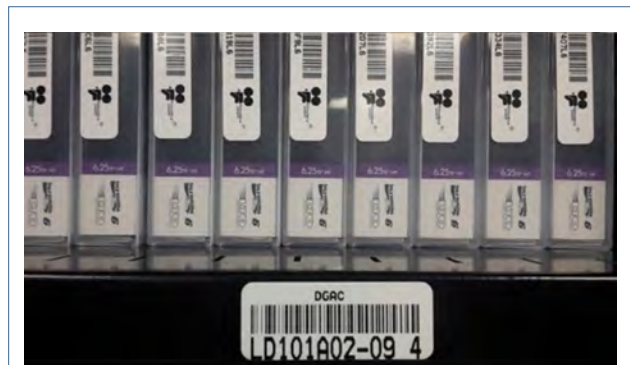


Figura 2. Acervo digital en LTO.
Fuente: elaboración propia.

El identificador de cada nicho en código de barras (ver figura 2) está organizado a modo de coordenadas espaciales estructurado con las siguientes componentes:

bóveda+nivel+número de anaquel+cara del anaquel+número de columna del anaquel+guión separador+número de entrepaño+espacio separador+dígito verificador

En donde el “+” significa “seguido de”

Al colocar contenedores en algún nicho de la estantería⁴, el operador establece la relación entre los CBs que identifican los contenedores y el CB del nicho en que se depositaron. Para estas tareas, se utiliza una aplicación ejecutada en un dispositivo móvil con un lector de CB integrado; misma aplicación que facilita la recolección posterior de contenedores. El método también facilita

³ También soporta la migración paulatina de sus más de 60,000 ejemplares fílmicos registrados en un catálogo iniciado en la década de 1990. Desde 2018, cada que se accede a un ejemplar fílmico este se revisa, se identifican sus latas con CBs, se migran sus metadatos de estado físico a CLAF y se da de baja del antiguo catálogo.

⁴ Estos identificadores también se utilizan para indicar talleres y laboratorios de la Filmoteca por donde los contenedores transitan temporalmente.

el trabajo eventual de reubicación y re-etiquetado de contenedores con *CBs* dañados o desprendidos⁵.

3. El contenedor de objetos digitales

Todo acervo de objetos digitales tiene que almacenarse, al menos por duplicado [7], en algún tipo de memoria cuyo soporte físico denominaremos **contenedor de objetos digitales (COD)**. El método de preservación que describiremos no depende ni del tipo de objeto ni de la tecnología del *COD*, por lo que convenimos en este artículo que el acervo digital se preserve en *CODs* como discos ópticos o de estado sólido, cintas o discos magnéticos, etc. Siendo la preservación una actividad continua [8], es indispensable revisar periódicamente los *CODs* y el estado de los objetos digitales que almacenan.

La preservación de un objeto digital debe prevenir o, cuando menos, detectar la más mínima alteración causada por el fenómeno *bit flip* [9]. La técnica para detectar cuando este fenómeno ya afectó⁶ a un objeto digital se conoce como *fixity checksum* [10] o “suma de comprobación de la fijeza”, a la que en lo subsecuente nos referiremos como “chequeo de fijeza” o simplemente “fijeza”. Esta técnica se apoya en algún algoritmo para calcular funciones de dispersión empleadas en criptografía [11]. El algoritmo debe de aceptar como entrada cualquier secuencia ordenada de *bits*, —como son los archivos de objetos digitales—, sin importar su tamaño ni lo que representa. El resultado calculado por el algoritmo será otra secuencia pequeña de *bits* conocida como *digest*, o *checksum* o *hash*. El quid del uso del algoritmo para checar la fijeza es que al repetir el cálculo sobre la misma entrada siempre obtendremos el mismo resultado, pero ante la misma entrada alterada en uno o más de sus *bits* obtendremos un *checksum* diferente.

Explicaremos dos *scripts* que emplean el paquete *FCIV* [12] con *MD5*⁷. Un *script* se usa para hacer **copiado**

íntegro a *CODs* de preservación. El otro se utiliza periódicamente para revisar los contenidos digitales de cada contenedor y el estado físico de los *CODs*.

Copiado íntegro

La figura 3 ilustra los tres pasos de un copiado íntegro. Después de efectuar los tres pasos de la figura 3, cuando la copia es íntegra, los archivos *fciv1.XML* y *fciv2.XML* serán idénticos, lo que se verifica comparándolos con un comando del sistema de archivos.

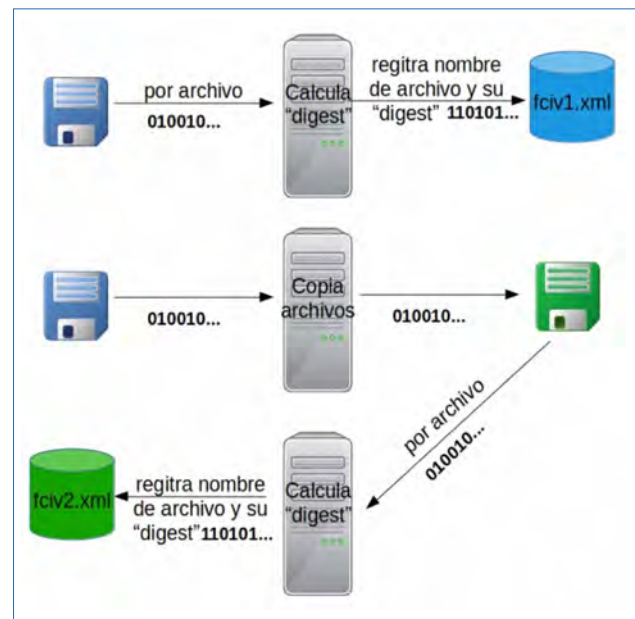


Figura 3. Flujo para verificar la integridad del copiado del COD azul al COD verde. Fuente: elaboración propia.

Por último, el *.XML* con los metadatos de fijeza se copia al COD que contiene los objetos digitales referenciados dentro del *.XML*. Esta es una solución práctica ya que almacenar los *.XML* por fuera de su *COD*, aumentará la complejidad del sistema y el trabajo de chequeo de la fijeza.

A modo de ejemplo, la figura 4 muestra un extracto del *.XML* resultante al aplicar nuestro *script* a los 86,400 archivos con fotografías *DPX* de una película digitalizada. En estos archivos *.XML*, cada par de etiquetas `<FILE_ENTRY></FILE_ENTRY>` delimita un par `<name></name>` con el nombre del archivo de un fotograma y otro `<MD5></MD5>` con su correspondiente *digest*, este último representado en “*base64*” y no en hexadecimal como es más usual [13].

⁵ En el caso del soporte filmico se adhieren dos etiquetas identificadoras, una en el lomo de la lata y la otra en una depresión de la tapa junto a otra etiqueta descriptiva del contenido.

⁶ No es posible anticipar las pérdidas por *bit flip*, de ahí la recomendación de copiar por triplicado.

⁷ *MD5* es menos demandante en recursos de cómputo que *SHA* cuyas cualidades de robustez criptográfica no interesan en la preservación.

```
<?XML version="1.0" encoding="utf-8"?>
<FCIV>
  <FILE_ENTRY>
    <name>MiArchivo_1.dpx</name>
    <MD5>bL/ZGbqnyeA8hHGUTY+LsA==</MD5>
  </FILE_ENTRY>
  <FILE_ENTRY>
    <name>MiArchivo_2.dpx</name>
    <MD5>x9-yeA8hHGUTY-9RAyOtGbqn</MD5>
  </FILE_ENTRY>
  :
  :
  <FILE_ENTRY>
    <name>MiArchivo_86400.dpx</name>
    <MD5>9eQu8hHGUTYp86cORhOP46hJ</MD5>
  </FILE_ENTRY>
</FCIV>
```

Figura 4. Extracto de un .XML para el copiado íntegro y el monitoreo de la fijeza.

El *script* para el copiado íntegro acepta estructuras arborescentes raíz-subcarpetas-archivos; la ruta o *path* de cada archivo se incluye como prefijo de cada nombre (omitido en el ejemplo). La verificación de integridad informa cuando la comparación de los .XML difiere, en cuyo caso el copiado debe reintentarse.

Revisión periódica de metadatos de fijeza

El *script* para la verificación periódica de la fijeza se apoya en el mismo principio que el copiado íntegro (ver figura 3), salvo porque en su primer y único paso trabaja con el contenido del COD verde para producir un nuevo .XML que se compara contra el .XML almacenado en el COD.

La ejecución del script detectará cualquier daño, ya sea electromecánico del COD o por *bit flip*.

Ante los daños, se tiene la esperanza de recuperar documentos dañados de la segunda copia que también debe ser sujeta a la verificación periódica de la fijeza de su contenido, —si bien la recomendación del NDSA [7] es copiar por triplicado—.

Necesariamente, la verificación periódica de la fijeza incluye la revisión del estado físico del COD, esto es que pueda leerse sin problemas.

Convenciones de estructura de CODs en la Filmoteca

Haciendo nuevamente referencia a la figura 3, el COD azul corresponde con almacenamiento masivo en NAS/SAN utilizado para depositar temporalmente los trabajos de digitalización y restauración digital; el COD

verde corresponde a cintas magnéticas LTO-6. Después de obtener dos copias en LTO-6 con sus metadatos de fijeza, la fuente se elimina del NAS/SAN para reutilizar el espacio de almacenamiento en nuevos proyectos de digitalización, restauración, conformación, cambios de formato y acceso a ejemplares digitales del acervo.

Al digitalizar una película y producir ejemplares en diferentes formatos, los objetos digitales resultantes se organizan en lo que denominamos una **unidad de respaldo**. La unidad de respaldo es el árbol de carpetas, subcarpetas y archivos que penden jerárquicamente de una raíz. Cada raíz corresponde a un proyecto de digitalización y recibe un nombre único compuesto por un folio seguido del nombre del título que CLAF registra y administra; sus subcarpetas reciben nombres de acuerdo con su contenido y convenciones adoptadas como “<número de rollo>”, “DPX”, “WAV”, “DCP”, “Full HD”, “h.264”, etc. Esta organización se respeta en los copiados íntegros a LTO y es recuperable.

4. Almacenamiento jerárquico

En esta sección se describen las operaciones para acceder a los ejemplares digitales y se establece el vocabulario para aclarar el funcionamiento del almacenamiento en niveles (*tiered storage*). Después, se exponen conceptos de almacenamiento jerárquico en tres niveles, su arquitectura y las componentes de una librería de cintas. Concluimos enumerando sus beneficios.

Acceso a los contenidos del acervo

Usando CLAF, cuando un usuario quiere acceder a algún ejemplar del acervo, tiene la opción de visionar un *proxy* de baja resolución (380x280 pixeles en formato h.264)⁸ de manera inmediata bajo video en demanda y puede solicitar otro ejemplar del mismo título en más alta resolución como *Full-HD* o *DCP*. Ante solicitudes de alta resolución, puesto que nuestro almacenamiento no permite mantener todo el acervo en discos, el ejemplar solicitado tiene que obtenerse de alguna LTO. Entonces, quien atiende la solicitud toma de su nicho la cinta LTO con el ejemplar solicitado, monta la cinta en una unidad de *i/o* de LTO y copia el ejemplar al almacenamiento en discos del NAS/SAN. Aunque el ejemplar ya

⁸ Este *proxy* invariablemente se genera en la unidad de respaldo de todo proyecto de digitalización y/o restauración. Al concluir el proyecto, su *proxy* se copia a otra unidad lógica de 4TB dedicada exclusivamente a soportar el visionado en línea de los títulos en soporte digital.

en discos puede visualizarse en el laboratorio LCRD, lo común es prestar la copia al solicitante para proyectarla en algún lugar como salas de cine digital o por televisión, incluso, publicarla en video en demanda por la Web si el formato del ejemplar es el adecuado. Nuevamente, la copia del ejemplar en discos eventualmente se eliminará para reutilizar el espacio.

Almacenamiento en dos niveles

Se dice que un ejemplar **está en línea** cuando existe una copia en disco y en ese momento puede reproducirse; en nuestro caso el disco corresponde al almacenamiento NAS/SAN. Decimos que un ejemplar **está fuera de línea** cuando el material está en cinta o genéricamente en un COD que carece de prestaciones para reproducir sus ejemplares de manera inmediata y en tiempo real (ver figura 5).

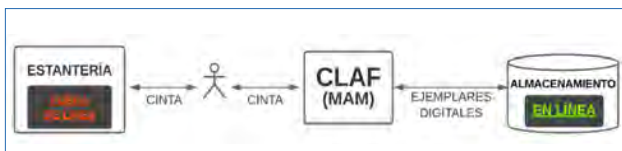


Figura 5. Almacenamiento en dos niveles.

Fuente: elaboración propia.

Almacenamiento jerárquico en tres niveles

Se dice que un ejemplar **está cerca de línea**, cuando la LTO que lo contiene está bajo el control de una librería de cintas como la que describiremos (véase la figura 6). Las librerías de cintas son un componente para integrar un **almacenamiento jerárquico en tres niveles**: en línea, cerca de línea y fuera de línea [14]. Si el lector utiliza otro tipo de COD para preservación como discos



Figura 6. Arquitectura del almacenamiento en tres niveles.

La estantería puede omitirse si la librería tiene capacidad suficiente. El ideal es mantener por triplicado librerías espejo remotas. Fuente: elaboración propia.

ópticos o magneto-ópticos, estos mismos niveles pueden soportarse con rocolas o jukeboxes de discos ópticos o magneto-ópticos en lugar de una librería de cintas. Ambas memorias jerárquicas son lógicamente equivalentes.

En la figura 6 se muestra una posible arquitectura de almacenamiento jerárquico en tres niveles:

- La librería de cintas monta o desmonta en unidades de i/o las cintas LTO bajo su control; sus contenidos están cerca de línea.
- Un sistema MAM similar a CLAF.
- Otro sistema que denominamos agente de servicios de la librería para el MAM.
- El almacenamiento de rápido i/o, en donde temporalmente residen copias de los ejemplares en línea.
- Estantería para cintas cuyos contenidos están fuera de línea.

Entre los servicios que el agente ofrece al MAM, dos son básicos: el de lectura que pone en línea ejemplares digitales solicitados, y otro de escritura para copiar íntegramente a LTO unidades de respaldo con metadatos de fijeza e informar al MAM en cuál cinta copió cada unidad de respaldo. El agente también administra la reutilización del espacio en línea.

La fotografía en la figura 7 muestra el interior de una librería de cintas típica. La librería de cintas consta de:

- Una colección de nichos que aceptan solo una cinta LTO por nicho. Cada cinta tiene su CB que la identifica de manera única.

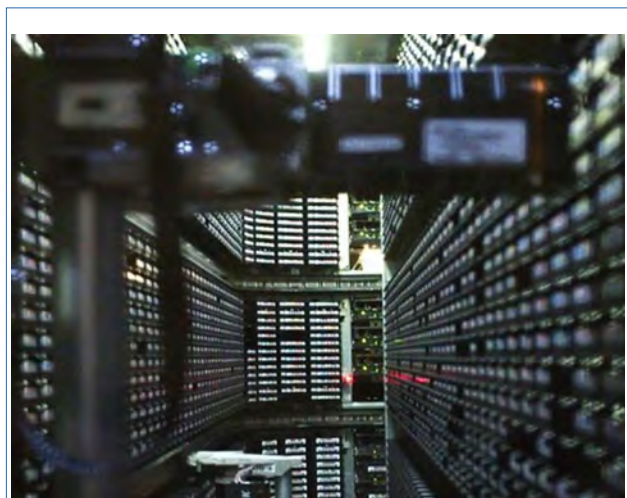


Figura 7. Fotografía del interior de una librería, cortesía de Derrick Coetzee (CCO 1.0).

Fuente: <https://bit.ly/3OfY06>.

- Drives de *i/o* de *LTO*, para una o más generaciones *LTO*, (no visibles en la fotografía).
- Uno o más brazos robóticos (al frente en la fotografía). Cada brazo “conoce” las coordenadas espaciales de cada cinta en sus nichos; coloca o retira *LTO*s de los nichos y de las unidades de *i/o*, de acuerdo con su generación *LTO* indicada en su *CB*.
- Un sistema que administra procedimientos de alimentación, movimiento y retiro de cintas, que “conoce” el identificador y la ubicación de cada *LTO* bajo su control.

Aplicaciones del almacenamiento jerárquico en tres niveles

Este almacenamiento jerárquico puede automatizar tareas de acceso y preservación con el beneficio de eliminar la posibilidad de omisión o error por su ejecución manual como hoy se hace en la Filmoteca. Las tareas cotidianas que una librería de cintas puede automatizar son:

- El copiado íntegro a *LTO* de nuevos objetos digitales cuando ingresan al acervo.
- La puesta en línea de ejemplares solicitados, lo que aminora los tiempos para su acceso.
- La automatización del chequeo periódico de la fijeza, en beneficio de la preservación del acervo. Proceso que puede detectar e informar sobre ejemplares dañados, e incluso apoyar y tomar acciones para eliminar los daños.

Además, y para enfrentar el continuo cambio tecnológico —que afecta y encarece la preservación del soporte digital—, el almacenamiento en tres niveles puede crecer para mantener cerca de línea cantidades gigantescas (del orden de *exabytes* [14]) y ayudar para controlar otros procesos eventuales e ineludibles como:

- Renovar el soporte físico (*COD*) o sustituir aquellos que se detecten dañados.
- Migrar todo un acervo copiándolo de manera íntegra a *CODs* de diseño más reciente y económico. El caso particular de la evolución de generaciones *LTO* es interesante [6].
- Transcodificar objetos digitales a nuevos formatos para la preservación de su reproducibilidad.

5. Conclusiones

La preservación y el acceso a los ejemplares digitales de imágenes en movimiento de una filmoteca similar a la de la UNAM, demandan enormes cantidades de almacenamiento para su preservación y hacen incosteable su

almacenamiento en línea o en la nube [15]. Desde que iniciamos la digitalización del acervo, en la Filmoteca elegimos las cintas *LTO* pues el costo por *byte* almacenado es el más bajo entre las múltiples opciones tecnológicas, además de que son un medio estable para conservar la fijeza de los datos por largos períodos de tiempo.

Las dos copias del acervo digital de la Filmoteca de la UNAM, hoy ocupan más de un *petabyte* en cintas *LTO* y han sido gestionadas desde 2015 con su sistema *CLAF*. Los procesos de acceso y chequeo de la fijeza requieren intervención manual que los hace susceptibles al error o la omisión. Las dimensiones ya alcanzadas de este acervo justifican plenamente la integración de una librería de cintas *LTO* para automatizar esos procesos en almacenamiento jerárquico en tres niveles.

En el artículo se describió cómo se organiza en *CLAF* el acervo digital, preparándolo para una relativamente fácil integración de una librería de cintas. La descripción se hizo de manera que los principios, arquitectura y técnicas empleadas puedan ser adaptadas a otros ambientes en que se utilice almacenamiento fuera de línea en contenedores digitales diferentes a las *LTO*.

Una arquitectura jerárquica de almacenamiento en tres niveles ofrece la posibilidad de automatizar: 1) el copiado íntegro de nuevos contenidos, 2) las sumas de comprobación de la fijeza periódicas a todo el acervo, y 3) el acceso a los ejemplares del acervo. Además, puede aprovecharse para controlar una eventual e ineludible migración o renovación de contenedores de objetos digitales, y para apoyar potenciales cambios de formato en respuesta al continuo cambio tecnológico. Finalmente, debe ser claro que la introducción de esta arquitectura a cualquier filmoteca o acervo digital de grandes dimensiones aumentará la confiabilidad y el horizonte de tiempo de preservación y acceso, optimizando costos de infraestructura tecnológica.

Agradecimientos

Los autores agradecen a Guadalupe Ferrer Andrade el haber depositado su confianza para integrar el *LCRD* y desarrollar *CLAF*, y a Hugo Villa Smythe, actual Director de la Filmoteca de la UNAM, por dar continuidad a estos proyectos. También es necesario reconocer la dedicación y calidad del trabajo de Genaro Pantoja Hernández y de Ever Iván López López, quienes cotidianamente y por varios años han digitalizado películas en el *LCRD*.

BIBLIOGRAFÍA

- [1] León Lastra G., Manuel Comi, Gustavo Lucio y Luis F. Maciel, “Sistematización del Acervo Fílmico de la Filmoteca de la UNAM”, 2017. En Rodríguez, Perla y Munera Fernández. *Conectando los saberes de bibliotecas, archivos y museos (BAM) en torno a la preservación de documentos analógicos y de origen digital*, pp. 153-172, IIBI-UNAM 2019, México, ISBN 978-607-30-1543-1.
- [2] Ultrium LTO, <https://www.lto.org/> [Consultado en marzo 30, 2023].
- [3] Fairbairn, Natasha, et al. “The FIAF Moving Image Cataloguing Manual”. L. Tadic y N. Goldman (eds.), *FIAF Cataloguing and Documentation Commission*, 2016. Disponible en agosto de 2017 en <https://www.fiafnet.org/pages/E-Resources/Cataloguing-Manual.html> [Consultado en marzo 30, 2023].
- [4] León Lastra G., “Situación actual y futura de la preservación del acervo digital de la Filmoteca de la UNAM”, 2021. En Rodríguez, Perla O. coordinadora, Creadores de memoria. *Los archivos sonoros y audiovisuales en México*, pp. 123-141, IIBI-UNAM 2021, México. Disponible en: <http://sistemas.iibi.unam.mx/publica20/conmutarl.php?arch=1&idx=394> [Consultado en marzo 30, 2023].
- [5] IBM “LTO Ultrium Cartridge Label Specification (Revision 6) Part Number 19P0034 EC - M10321”, Feb., 2009. Disponible en <https://www.ibm.com/support/pages/ibm-lto-ultrium-cartridge-label-specification> [Consultado en marzo 30, 2023].
- [6] Ultrium LTO, “Roadmap”, 2023. Disponible en <https://www.lto.org/roadmap/> [Consultado en marzo 30, 2023].
- [7] National Digital Stewardship Alliance (NDSA), “Levels of Digital Preservation”, versión 2, 2019. Disponible en <https://nds.org/publications/levels-of-digital-preservation/> [Consultado en marzo 30, 2023].
- [8] Trevor Owens, “La teoría y el oficio de la preservación digital”. Coord. Isabel Galina Russell. Trad. Pedro Ángeles Jiménez et al., 2022, *Instituto de Investigaciones Bibliográficas-UNAM*. ISBN: 978-607-30-7031-7.
- [9] Robert Elder, “What causes bit flips in computer memory”, 2023. Disponible en <https://blog.robertelder.org/causes-of-bit-flips-in-computer-memory/> [Consultado en marzo 30, 2023].
- [10] Consultative Committee for Space Data Systems (CCSDS), “Reference Model for an Open Archival Information System (OAIS)”, 2012. *Recommended Practice, CCSDS 650.0-M-2, Magenta Book. Management Council of the Consultative Committee for Space Data Systems, USA*. Disponible en <https://public.ccsds.org/pubs/650x0m2.pdf> [Consultado en marzo 30, 2023].
- [11] National Digital Stewardship Alliance (NDSA), “Checking your digital content”, 2014. Disponible en <https://www.digitalpreservation.gov/documents/NDSA-Fixity-Guidance-Report-final100214.pdf> [Consultado en marzo 30, 2023].
- [12] Internet Archive, “Files for FCIV_v2_05”, https://archive.org/download/FCIV_v2_05 [Consultado en febrero 6, 2024].
- [13] Malte Hillmann, “fcivparser”, 2017. Disponible en <https://github.com/MalteHillmann/fcivparser/blob/master/md5.xml> [Consultado en marzo 30, 2023].
- [14] Wikipedia, “Memory hierarchy”, 2023. Disponible en https://en.wikipedia.org/wiki/Memory_hierarchy [Consultado en marzo 30, 2023].

[15] Coordinación de Nuevas Tecnologías e Informática (CoNTI), “Reporte sobre costos para mantener una copia espejo en la nube del acervo digital de la Filmoteca de la UNAM”. *Reporte interno de la Coordinación de Nuevas Tecnologías e Informática para la Dirección General de Actividades Cinematográficas, UNAM*, noviembre de 2018.

Fecha de recepción: 28 de abril de 2023

Fecha de publicación: febrero de 2024

Cómo se cita

G. León Lastra, G. Lucio José, M. Comi Xolot, L. F. Maciel Mercado. “Preservación y acceso a objetos digitales fuera de línea: experiencias en la Filmoteca de la UNAM”, *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].