Acesso Aberto Angola

Recomendaciones de preservación digital para el Repositorio Nacional de Angola

Trabajo derivado del proyecto de colaboración entre Redalyc UAEM, AmeliCA, MESCTI y U. Óscar Ribas financiado por UNESCO para el desarrollo de una estrategia de Acceso Abierto en Angola

Coordinadores

Arianna Becerril-García, Eduardo Aguado-López, Alejandro Macedo-García, Eurico Wongo-Gungula

Marzo, 2022
Vol. 33

unesco · Governo de Angola · mescti.gov.ao · UAEM · redalyc.org UAEM · AmeliCA · UÓR Universidade Óscar Ribas

Recomendaciones de preservación digital para el Repositorio Nacional de Angola. Trabajo derivado del proyecto de colaboración entre Redalyc UAEM, AmeliCA, MESCTI y U. Óscar Ribas financiado por UNESCO para el desarrollo de una estrategia de Acceso Abierto en Angola. Vol. 33.

Coordinadores: Arianna Becerril-García, Eduardo Aguado-López, Alejandro Macedo-García, Eurico Wongo-Gungula.

Marzo, 2022

Citación sugerida: United Nations Educational, Scientific and Cultural Organization, Ministério do Ensino Superior, Ciência, Tecnologia e Inovação de Angola, Redalyc-Universidad Autónoma del Estado de México, AmeliCA & Universidade Óscar Ribas (2022). Recomendaciones de preservación digital para el Repositorio Nacional de Angola (Vol. 33). Zenodo. https://doi.org/10.5281/zenodo.6347550

El presente escrito se desprende del trabajo realizado en 2021 como parte del proyecto de colaboración entre la UNESCO y el Sistema de Información Científica Redalyc. El proyecto contó con la colaboración del Gobierno de Angola a partir del Ministério do Ensino Superior, Ciência, Tecnologia e Inovação (MESCTI), la Universidad Autónoma del Estado de México (México), AmeliCA y la Universidade Óscar Ribas (Angola), y tuvo como objetivo realizar un diagnóstico integral de la comunicación científica de Angola, una hoja de ruta crítica para el desarrollo del Acceso Abierto en el país y una serie de desarrollos tecnológicos en favor de una comunicación científica en abierto no comercial. El proyecto derivó en un conjunto de 38 trabajos que se hacen públicos con el objetivo de aportar e incidir en una infraestructura tecnológica, legal y de conocimiento de Acceso Abierto no comercial en Angola.

Coordinadores:

Arianna Becerril-García, Eduardo Aguado-López, Alejandro Macedo-García, Eurico Wongo-Gungula.

Equipos de trabajo:

México
Sistema de Información Científica Redalyc
Universidad Autónoma del Estado de México

Investigación:
- Sheila Godínez-Larios
- Liliana González-Morales
- Marco Antonio Estrada-Medina
- Pedro Villegas-Hernández
- Brenda Uribe-Martínez

Desarrollo de software:
- Alma Rosa Segundo Escobar
- Daniel Mejía Antolín
- Edgar Juárez Escamilla
- Thania del Carmen Colín Álvarez
- Jonatan E. Montes de Oca Ríos
- Jorge Juan Díaz Carbajal
- Domingo Anzaldo Bibiano

Arte y diseño:
- Bernardo Bernal-Gómez
- Priscila Dávila-Morales
- Nayeli Lara-Martínez
- Abril Carmona Ochoa

Diagnóstico editorial:
- Lidia Abigail Villagómez Beltrán
- Ana Lilia Aladín Díaz
- Daniel Francisco Flores García

Comunicación:
- Pamela Amarillas Nava

Traducción:
- Cristell Rubí Hernández Cruzaley
- Jessica Mireya Trujillo Zúñiga
- Martha Paulina Ibarra Quintana
- Nancy Verónica Derbéz Cruz
- Paola Andrea Carbajal García

Angola
Universidade Óscar Ribas

Investigación:
- Carla Santana
- Josefina Castillero
- Inês Portugal

Tabla de contenido

I. PRESERVACIÓN DIGITAL 6

AMENAZAS FÍSICAS 6

OBSOLESCENCIA 7

OBSOLESCENCIA DE ARCHIVOS 7

SUGERENCIAS PARA EVITAR LA OBSOLESCENCIA DE ARCHIVOS 7

OBSOLESCENCIA DE EQUIPOS Y HARDWARE 7

SUGERENCIAS PARA EVITAR LA OBSOLESCENCIA DE EQUIPOS Y HARDWARE 8

II. ESTRATEGIAS DE PRESERVACIÓN DIGITAL 8

RESPALDO 8

RENOVACIÓN DEL SOPORTE DE ALMACENAMIENTO 9

REPLICACIÓN 9

UTILIZACIÓN DE MEDIOS PERSISTENTES 9

MANTENIMIENTO FÍSICO DE LOS EQUIPOS 9

ARQUEOLOGÍA DIGITAL 10

MIGRACIÓN 10

USO DE ESTÁNDARES 10

NORMALIZACIÓN 11

CONTROL DE LA AUTENTICIDAD 11

CONTROL DE LA ESTABILIDAD O INTEGRIDAD 12

EMULACIÓN 12

METADATOS DE PRESERVACIÓN 13

III. MODELO DE REFERENCIA OAIS (OPEN ARCHIVAL INFORMATION SYSTEM) 14

DSPACE Y LA PRESERVACIÓN DIGITAL 15

METADATOS DE PRESERVACIÓN 17

PREMIS 17

METS 18

JHOVE: extracción automática de metadatos *18*

IV. PRESERVACIÓN DIGITAL EN REPOSITORIOS **19**

CASOS DE ESTUDIO DE PRESERVACIÓN DIGITAL EN REPOSITORIOS NACIONALES E INSTITUCIONALES ... 20

Repositorio UNAM *20*

Wheaton College Digital Repository *21*

MIT institutional repository *22*

V. RECOMENDACIONES DE PRESERVACIÓN DIGITAL PARA EL REPOSITORIO DE ANGOLA **23**

VI. CONCLUSIÓN **25**

BIBLIOGRAFÍA **26**

Figuras

Figura 1 Modelo OAIS (CCSDS Secretariat, 2002) 15
Figura 2 Proceso de incorporación de DSpace (Chapter 2. DSpace System Documentation, 2002) 16
Figura 3 Modelo de datos PREMIS (PREMIS Working Group, 2008) 17
Figura 4 Arquitectura para depósito (Ramírez Molina, Ana Yuri, 2021) 21
Figura 5 Flujo de trabajo para la administración de contenido digital (MIT Libraries, 2018) 23

I. Preservación digital

Según la UNESCO, la preservación digital consiste en los procesos destinados a garantizar la accesibilidad permanente de los objetos digitales. Para lograrlo, es necesario buscar maneras de representar lo que se había presentado originalmente a los usuarios haciendo uso de equipos y programas informáticos que permiten procesar los datos.

Los documentos digitales han demostrado ser mucho más frágiles desde el punto de vista de la conservación, pues en ella se involucra una serie de factores que no aplica en los documentos no digitales, como la dependencia de un equipo que interprete el contenido, el cambio constante de los dispositivos de almacenamiento y lectura que dejan obsoletos a los anteriores, las modificaciones que se pueden ir agregando en el mismo documento, etc.

Los documentos no digitales están sujetos a ciertas amenazas (pérdida, derrame de líquido, incendio, destrucción, desgaste por el paso del tiempo, etc.). Lo mismo sucede con los documentos digitales, descontando las cuestiones legales y económicas, se pueden perder de dos maneras: por daño físico u obsolescencia, o por una combinación de ambas.

Amenazas físicas

Pueden ocurrir en cualquier momento por razones internas o externas y afectar a los materiales necesarios para acceder al contenido digital.

Los daños ocurren de diversas maneras:

- Desastre natural (incendio, terremoto, tormenta, etc.).
- Condiciones inadecuadas de almacenaje, como temperaturas y porcentajes de humedad relativa.
- Fallos de energía.
- Mal mantenimiento de los equipos de hardware.
- Abuso en la utilización de los mecanismos manuales.
- Fallos humanos, como la incorrecta manipulación de los medios, el derramamiento de líquidos, caídas de equipos, etc.
- Actos malintencionados, como robos, virus, sabotajes, etc.
- Deterioro natural de los equipos.

Obsolescencia

La amenaza de pérdida por obsolescencia es aún más grave que la de deterioro físico, puesto que es más difícil de controlar al haber más factores involucrados.

Obsolescencia de archivos

Los motivos por los que un archivo podría quedar obsoleto son los siguientes:

- El software de lectura está descontinuado.
- Un software moderno no lee formatos antiguos.
- Un formato se hace más complejo o es reemplazado.
- El formato no es lo suficientemente masivo, por lo que desaparecen las aplicaciones compatibles.

Sugerencias para evitar la obsolescencia de archivos

Se deben priorizar los objetos creados con aplicaciones obsoletas o con peligro de obsolescencia, revisando previamente si el contenido de estos documentos se debe preservar. Para elegir el nuevo formato de la información, se debe optar por aquellos que tengan más inmunidad a la obsolescencia, por ejemplo:

- Que tengan buena compatibilidad hacia atrás.
- Que sean ampliamente utilizados globalmente.
- Que no sean excesivamente complejos ni tampoco demasiado simples.
- Que tenga chequeo de errores incorporado.
- Que tengan un ciclo de actualizaciones relativamente frecuente.
- De preferencia, que sean formatos abiertos.

Obsolescencia de equipos y hardware

En las últimas décadas muchos equipos están diseñados para tener un periodo de vida útil relativamente corto. También las nuevas versiones de software, que generalmente otorgan posibilidades más amplias, van exigiendo mayor efectividad de hardware, lo que provoca que ambos crecimientos y obsolescencias sean proporcionales.

Además, tecnologías externas como las conexiones periféricas (por ejemplo, los puertos USB han sustituido a tecnologías de conexión como RS-232) o los dispositivos externos como pendrives, discos duros externos, escáneres e impresoras también han ido evolucionando.

Sugerencias para evitar la obsolescencia de equipos y hardware

- Comprar medios de calidad, aunque esto resulte más costoso.
- No proyectar la vida útil de los equipos a más de cinco años.
- No realizar compras exageradas, considerando que en algunos años habrá que renovar y algunos equipos podrían quedar sin uso.
- Realizar pruebas periódicas para ver el estado de los datos.

II. Estrategias de preservación digital

Sabemos que ningún soporte de almacenamiento puede durar para siempre, por lo que hay que ir constantemente buscando las mejores opciones para mantener la legibilidad de los archivos. Existen distintas estrategias para preservar los datos, pero ninguna por sí sola es capaz de solucionar el problema. Por el momento, la mejor práctica, es la mezcla de éstas.

Respaldo

Se trata simplemente de hacer un duplicado exacto del objeto que se intenta preservar. En cualquier programa de preservación digital, el respaldo debe ser considerado el recurso mínimo de mantenimiento, que debe abarcar a la totalidad de los archivos, incluso a aquellos que son considerados de bajo valor.

La duplicación es un componente esencial de la preservación puesto que se ocupa de la pérdida de datos por problemas de hardware u otros fallos como el mal funcionamiento de los equipos, el deterioro, desastres naturales, etc.

Sin embargo, al no hacer frente al problema de la obsolescencia, el respaldo tan solo es considerado una estrategia a corto plazo. El respaldo debe siempre ir acompañado de almacenamiento remoto, de esta manera se evita que un mismo desastre altere la totalidad de las copias.

Renovación del soporte de almacenamiento

También es una estrategia para reducir el riesgo de pérdida debido al deterioro de los soportes, y que se basa en realizar una copia, sin alterar en absoluto la información digital, pero con la diferencia de que se buscan soportes de almacenamientos más modernos. Por ejemplo, el traspaso de datos desde un disco duro antiguo a uno nuevo o a un SSD.

Replicación

Es el duplicado y copiado de la información en uno o más sistemas, y su principal fortaleza es el almacenamiento en más de un lugar para, de esta manera, evitar que una misma alteración, intencional o accidental, o el mismo desastre natural pueda hacer que se pierda la totalidad de la información.

Es importante señalar que mientras más copias almacenadas existan, mucho mayor es el costo de la mantención, el respaldo y la actualización periódica que requiere todo el archivo, pero sin duda, ningún programa de preservación digital puede considerarse como tal, si no tiene al menos un depósito físico remoto.

Utilización de medios persistentes

Se trata de la utilización de medios más resistentes y perdurables. Ayuda de esta manera a reducir las pérdidas ocasionadas por el deterioro de los medios de almacenamiento comunes.

Ejemplos de soportes persistentes son los CDs o DVDs de platino o de oro. Sin embargo, no podemos confiar ciegamente en esto, puesto que nada puede hacer frente a amenazas tales como la obsolescencia de los medios de codificación o los desastres naturales.

Mantenimiento físico de los equipos

Para intentar evitar la pérdida de los dispositivos por motivos de deterioro, las sugerencias más comunes para todos los dispositivos son:

- Mantener estables las condiciones de temperatura y humedad.
- Controlar el polvo.
- Tener equipos de detección de fuego, humo y temperaturas extremas.
- Prohibir comer, beber y fumar en las dependencias donde se encuentren los medios.
- Para la manipulación, tener las manos siempre limpias y secas.

Arqueología digital

Se trata de una estrategia de emergencia que intenta rescatar los contenidos digitales que estaban almacenados en medios que han sido dañados físicamente, o que pertenecían a entornos de hardware y software obsoletos o dañados.

Para llevar a cabo esta estrategia, es necesario aplicar técnicas especializadas para recuperar la información que, aunque sigue estando almacenada, no puede ser interpretada por los medios y se ha convertido en ilegible.

Migración

La información digital es inútil si no está codificada para que pueda ser legible por personas, por lo tanto, el principal objetivo de la migración es conseguir la mantención de la accesibilidad a los recursos digitales para que cualquier usuario, en cualquier momento, pueda recuperarla sin que los cambios tecnológicos la alteren.

Esta estrategia se compone de una serie de tareas organizadas y diseñadas para lograr la transferencia periódica de información digital desde un sistema a otro más reciente, más seguro o que entregue mejores posibilidades. Esta transferencia puede ser de un formato a otro, de un sistema operativo a otro o de un lenguaje de programación a otro, para así se intentar asegurar por un tiempo más la accesibilidad al objeto digital.

Uso de estándares

Para la preservación digital se recomienda utilizar siempre estándares abiertos, para que la legibilidad de los documentos no esté condicionada por el devenir de la compañía fabricante del software o archivo.

El objetivo principal de los estándares abiertos es que se pueda crear la competencia entre diversos tipos de programas o archivos para así mejorar la calidad, y a la vez impedir que el vendedor ejerza el control sobre el futuro de la información.

Ejemplos de estándares abiertos:

- Software: PDF/X, PDF/A, OpenDocument, GIMP
- Hardware: ISA, PCI, AGP
- Sistemas operativos: LINUX
- Formatos: PDF, TXT, JPEG, PNG, Theora, FLAC, XML

Normalización

Como un depósito digital cuenta con muchos tipos de archivos digitales, se debe elegir un formato para cada grupo de archivos del mismo tipo, normalmente se optará por el que entregue mejores posibilidades de longevidad, funcionalidad y preservación, evitando así problemas de complejidad y coste.

La normalización pretende que la representación del contenido pueda ser ajena a aplicaciones informáticas específicas y que sea posible lograrla con algún software abierto que pueda ir cambiando fácilmente con los avances de la tecnología.

Control de la autenticidad

Esta estrategia está pensada como control de calidad de los documentos que serán preservados para asegurar la legitimidad de la información digital. El control de la autenticidad de los documentos digitales cobra vital importancia para todo tipo de archivos, en especial para aquellos que se utilizarán con fines legales, financieros o científicos.

Algunas medidas de control de autenticidad que un archivo debe tener son:

- Documentar la procedencia y la historia del documento digital.
- Utilización de metadatos de preservación que documenten la identidad y la integridad del objeto.
- Evaluación periódica y pruebas de autenticidad de sus documentos.
- Documentación de cada uno de los cambios que sufre un objeto digital.
- Utilización de marcas de agua o filigranas digitales que dificultan el copiado desautorizado.

Control de la estabilidad o integridad

Los riesgos que sufren la integridad y la autenticidad son los mismos, es decir, la facilidad con que un documento digital puede ser modificado. Existen muchas maneras de que esto suceda, por ejemplo:

- Errores humanos.
- Errores naturales que se producen en los sistemas de almacenamiento.
- Deterioro de los soportes.
- Virus informáticos.
- Modificaciones malintencionadas de hackers o cualquier persona que tenga acceso a los archivos.

Para intentar mantener y comprobar la integridad de los documentos digitales, existen diversas estrategias que se pueden complementar entre sí:

- Firmas digitales.
- Utilizar metadatos de preservación.
- Documentar cada una de las transformaciones que sufren los documentos.
- Sumas de verificación, por ejemplo, guardar el número de bits del archivo.

Este proceso se puede combinar con el algoritmo de reducción criptográfica MD5 (Message-Digest Algorithm 5) que permite verificar si los contenidos descargados desde Internet son fieles a sus originales.

Emulación

Esta estrategia combina elementos de software y de hardware para reproducir en un contexto distinto al original las características esenciales del archivo. La idea principal es que un formato antiguo u obsoleto funcione en un ambiente informático nuevo que originalmente no reconoce al viejo formato.

Para llevar a cabo la emulación se requiere de un nuevo software, un emulador, que traducirá los códigos e instrucciones desde el entorno computacional antiguo para que se ejecute correctamente en el nuevo. Para crear el emulador se utiliza la llamada “ingeniería inversa” del software original, que analiza todas las características de éste con la finalidad de determinar de qué está hecho y de qué manera funciona. Profundizando en este estudio del funcionamiento, se pueden entender, modificar y mejorar las características del software.

La emulación se ha concretado en varios proyectos, entregando a menudo resultados prometedores. Sin embargo, la técnica es un poco difícil de aplicar, puesto que alrededor de ella circulan una serie de cuestiones complejas, como las técnicas de creación de los emuladores, los pasos administrativos para montar las especificaciones, la documentación de los sistemas que se emularán y las problemáticas legales que significan obtener los derechos de propiedad intelectual del hardware o software en cuestión.

Metadatos de preservación

En los archivos digitales no es tan fácil asociar el contenido del documento con su soporte, por lo tanto, se recurre a los metadatos, que son elementos que acompañan al documento digital clasificándolo y describiéndolo, para así facilitar su identificación y recuperación.

Además de las funciones de clasificación y descripción, los metadatos aportan información específica para intentar recrear el entorno informático original, haciendo frente a la facilidad de alteración que todo documento electrónico tiene.

Una de las iniciativas más importantes para estandarizar los metadatos es, sin duda, Dublin Core, que como se menciona en *Estructuración de metadatos y textos completos (Dublin Core NISO Z39.85-2012, XML JATS NISO Z39.96-2019): recomendaciones para su implementación en el Repositorio Nacional de Angola*, es el estándar que se utilizará en el proyecto de Angola. Los metadatos ofrecen mejoras sustanciales que favorecen la preservación, pero no son metadatos destinados específicamente a ella, por lo que algunos han intentado agregar una cuarta categoría destinada a la preservación: datos descriptivos, para localizar; técnicos, para visualizar y utilizar; administrativos, para el control de la integridad y autenticidad; y legales, para evaluar las posibilidades y limitaciones de uso.

Los **metadatos de preservación** están destinados a almacenar los detalles técnicos sobre el formato, la estructura, el acceso y el uso de los contenidos digitales, la historia de todas las acciones realizadas en el recurso, incluyendo los cambios, la información de autenticidad, las características técnicas o la historia de la custodia y las responsabilidades y la información sobre los derechos con que se cuenta para realizar las acciones de preservación.

Los metadatos pueden estar encapsulados con el mismo documento, en un documento independiente, o en ambas partes. En espera de que se normalice la manera de incluir los metadatos, lo ideal es hacerlo con la última opción, pero esta manera de aplicar ambos procedimientos es necesario mantenerla bajo un exhaustivo control, puesto que cualquier error puede generar versiones diferentes en cada lugar.

III. Modelo de referencia OAIS (Open Archival Information System)

Se trata de una recomendación para el desarrollo de un amplio consenso sobre los requisitos que debe cumplir un archivo para preservar a largo plazo la información digital. Un sistema OAIS es un archivo que consiste en una organización de equipos humanos y sistemas que tienen la responsabilidad de preservar información y hacerla disponible para una comunidad específica. Un sistema estará de acuerdo con OAIS si soporta el modelo de información descrito en la norma, que no especifica ningún método de implantación.

OAIS propone seis pasos ineludibles para un programa, cuatro de ellos son actividades a las que se somete el material almacenado:

1. Incorporación (*ingest*)
2. Almacenamiento (*archival storage*)
3. Gestión (*data management*)
4. Acceso/difusión (*access/disemination*)

Otras dos hablan del funcionamiento del depósito:

5. Planificación para la preservación (*preservation planning*)
6. Gestión del depósito (*archive administration*)

Dentro de esta terminología propuesta, se encuentra el paquete de información IP (Information Package), que se refiere al conjunto que conforma el objeto digital con sus metadatos. El IP tiene tres versiones según su estado:

- **SIP (Submission Information Package):** es el paquete que entra al depósito.
- **AIP (Archival Information Package):** es el paquete que ya tiene las modificaciones necesarias para ser almacenado en *bitstream* o cadena de bit.
- **DIP (Dissemination Information Package):** es el paquete listo para ser puesto a disposición de los usuarios.

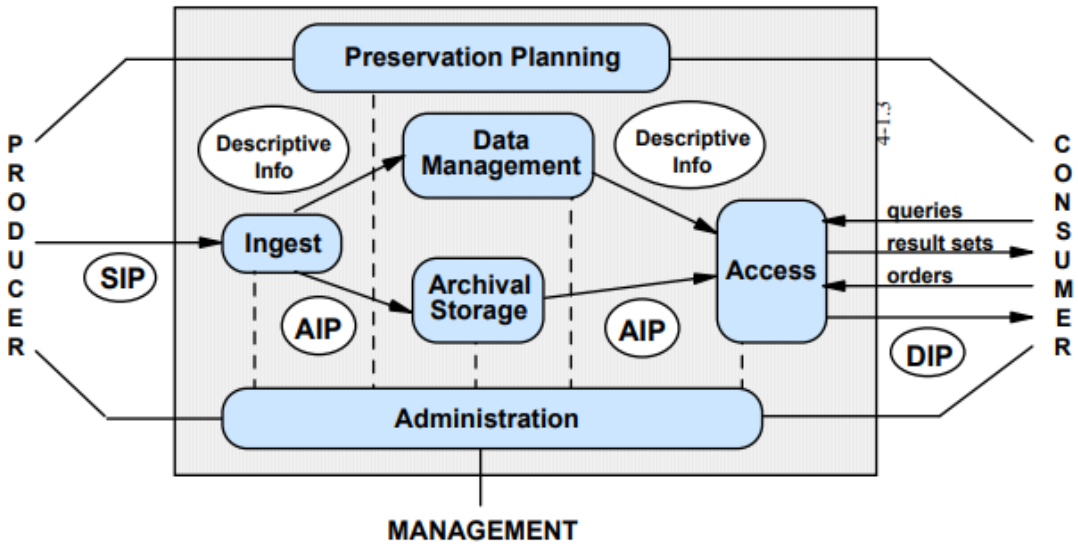

Figura 1 Modelo OAIS (CCSDS Secretariat, 2002)

En el DIP, posiblemente la información estará muy cambiada a como entró en el SIP, pues puede tener nuevos metadatos o haber sufrido migraciones. Cuando los materiales ingresan al programa, se deben tomar ciertas medidas de reconocimiento, como el formato, las especificaciones de éste, ver si dentro de este formato hay otros involucrados, etc. El modelo también propone que si el ingreso lo hace otra institución, ésta debe tener las responsabilidades de preparar el material para que tenga las condiciones propuestas.

DSpace y la preservación digital

Cualquier repositorio es, en parte, producto del software de gestión que se esté utilizando. La estructura, los formatos, los registros y hasta la preservación van a estar condicionados por el software que se utilice para gestionarlo. En este caso, se hablará del software DSpace.

DSpace en sí no garantiza la conservación de sus materiales digitales; sin embargo, está diseñado para desempeñar un papel central en la estrategia general de preservación digital, cumpliendo con el modelo OAIS visto previamente. Proporciona las funciones SIP, AIP y DIP. Usa METS como contenedor AIP (Archival Information Package).

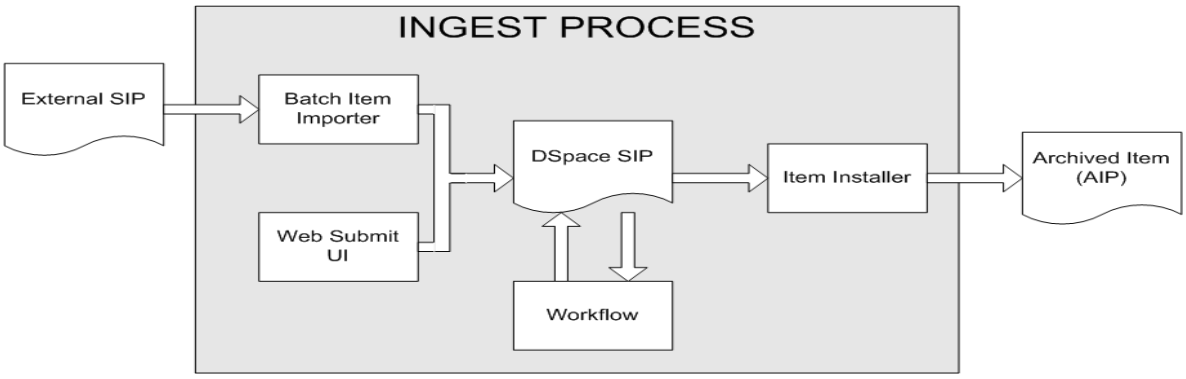

Figura 2 Proceso de incorporación de DSpace (Chapter 2. DSpace System Documentation, 2002)

DSpace identifica dos niveles de preservación digital:

- Preservación de bits: garantiza que un archivo permanezca exactamente igual a lo largo del tiempo (no se cambia ni un solo bit) mientras los medios físicos evolucionan a su alrededor.
- Preservación funcional: el archivo cambia con el tiempo, de modo que el material sigue siendo utilizable inmediatamente de la misma forma que lo era originalmente, mientras que los formatos digitales (y los medios físicos) evolucionan con el tiempo. Algunos formatos de archivo se pueden conservar funcionalmente mediante una migración de formato sencilla (por ejemplo, imágenes TIFF o documentos XML). Otros formatos son propietarios o, por otras razones, son mucho más difíciles de conservar funcionalmente.

Debido a que no es posible predecir los formatos que elegirán todos los usuarios para sus materiales de investigación, DSpace permite elegir tres niveles de conservación para un formato determinado: admitido, conocido o no admitido.

- Los formatos admitidos son aquellos que se cree que pueden conservarse funcionalmente utilizando técnicas de emulación o migración de formato. Los ejemplos incluyen TIFF, SGML, XML, AIFF y PDF.
- Los formatos conocidos son aquellos que no puede prometerse que van a preservarse, como los formatos patentados o binarios, pero que son tan populares que es probable que surjan herramientas de migración de terceros para ayudar con la migración de formatos. Los ejemplos incluyen Microsoft Word y Powerpoint, Lotus 1-2-3 y WordPerfect.
- Los formatos no admitidos son aquellos sobre los que no se conoce lo suficiente como para realizar algún tipo de preservación funcional. Esto incluiría algunos formatos propietarios o un programa de software único en su tipo.

DSpace proporciona algunos valores predeterminados para los formatos admitidos, conocidos y desconocidos. Se deben determinar los valores adecuados en función de la estrategia de conservación.

Metadatos de preservación

PREMIS

Este diccionario de datos es una traducción del modelo OAIS a unidades semánticas implementables, bajo la forma de un esquema de metadatos específicos para preservación, y sustentado en encuestas sobre sistemas reales de repositorios de preservación.

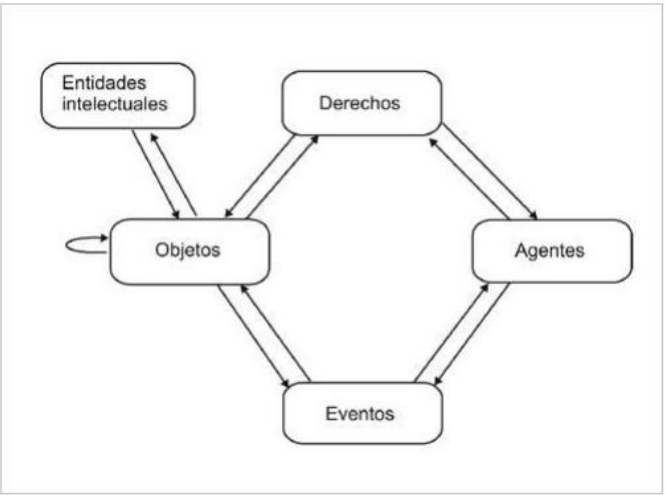

Figura 3 Modelo de datos PREMIS (PREMIS Working Group, 2008)

Las entidades en el modelo de datos PREMIS se definen de la siguiente manera:

- Entidad intelectual: conjunto de contenidos que se considera una única unidad intelectual a efectos de gestión y descripción, por ejemplo, un libro, un mapa, una fotografía o una base de datos. Una *entidad intelectual* puede comprender otras *entidades intelectuales*.
- Objeto [digital]: unidad discreta de información en formato digital.
- Evento: acción que al menos afecta a un *objeto* o *agente* asociado o conocido por el repositorio de preservación.
- Agente: persona, organización o programa/sistema informático asociado a los *eventos* durante la vida de un *objeto*, o a los *derechos* ligados a un *objeto*.

- Derechos: declaración de uno o varios derechos o permisos pertenecientes a un *objeto* o *agente*.

METS

Es un formato estándar para codificación y transmisión de metadatos. Está pensado principalmente para el envío de los ficheros, imágenes y objetos multimedia de una biblioteca digital. METS utiliza la estructura de etiquetas tipo XML. Un documento METS está compuesto por 7 secciones principales (no todas son obligatorias):

1. metsHdr (*METS Header*): información del documento METS: fecha y hora, nombre de la persona que lo crea, etc.
2. dmdSec (*Descriptive Metadata*): descripción del objeto al que se hace referencia en el documento METS.
3. admSec (*Administrative Metadata*): contiene los metadatos administrativos sobre los ficheros que forman el objeto digital y el material fuente del que se han obtenido dichos ficheros. Existen cuatro formas principales de metadatos administrativos disponibles para su utilización en un documento METS:
    a. Metadatos técnicos
    b. Metadatos sobre derechos y propiedad intelectual
    c. Metadatos sobre la fuente
    d. Metadatos sobre el origen digital
4. fileSec (*File groups*): son los ficheros que comprenden la versión electrónica del objeto digital.
5. structMap (*Structural Map*): define la estructura jerárquica del objeto y nos permite navegar por él.
6. smLink (*Structural Links*): se utiliza para indicar hipervínculos en el mapa estructural.
7. Behaviour Section: se utiliza para asociar comportamientos ejecutables con los contenidos del objeto METS.

JHOVE: extracción automática de metadatos

La extracción automática de metadatos se puede realizar con la herramienta JHOVE desarrollada por la Universidad de Harvard, que permite la identificación automática, validación y caracterización de un conjunto de tipos de objetos digitales.

Sería conveniente tener en cuenta, para una posible instalación, que DSpace ofrece complementos (*add-ons*) con una versión abreviada de JHOVE. Con esta herramienta, DSpace proporciona control del formato, de la extensión del archivo/*bitstream* y comprueba la presencia de virus.

JHOVE utiliza una arquitectura de complementos extensible que se puede configurar en el momento de su invocación para incluir cualquier módulo de formato específico y manejador de salida que se desee. La versión inicial de JHOVE incluye módulos para flujos de bytes arbitrarios, texto codificado en ASCII y UTF-8, GIF, JPEG2000 y JPEG, e imágenes TIFF, audio AIFF y WAVE, PDF, HTML y XML; y controladores de salida de texto y XML.

IV. Preservación digital en repositorios

El desarrollo de procesos de preservación para un repositorio digital confiable requerirá la integración de nuevos métodos, políticas, estándares y tecnologías. Los repositorios digitales deben poder conservar materiales electrónicos durante periodos al menos comparables a los métodos de conservación existentes.

En la mayoría de los repositorios se incluiría la preservación a largo plazo de materiales digitales como una función clave de éste, debido a que los repositorios deben ser un vehículo para hacer frente a las obligaciones de preservación de datos, pero contrario a esto, algunos repositorios están más preocupados por el acceso que por la preservación, pocos son los repositorios que tienen una política formal de preservación.

Los criterios sobre los recursos digitales según el Digital Preservation Handbook **(**Manual de Preservación Digital) son:

- Que la institución tenga pleno derecho a manipular los datos para asegurar su acceso en entornos informáticos del futuro.
- Que el recurso se encuentre en un formato legible y de probable funcionamiento en el futuro.
- Que el recurso esté en un soporte gestionable para su transferencia y/o almacenamiento.
- Que el recurso disponga de documentación, incluyendo los metadatos.

La selección de recursos para su preservación incluye la decisión sobre qué formatos, qué versiones, qué tipo de material adicional incluir y qué atributos se quieren preservar (datos y funcionalidad, apariencia y esencia). Un ejemplo de selección de tipo de formato es la Universidad de Loughborough, que utilizó los siguientes formatos de archivos:

- Texto: Microsoft Word, PDF
- Otros de Microsoft Office: PowerPoint, Excel, Access
- Archivos de video/animación: Flash, QuickTime
- Archivos basados en web: HTML
- Imágenes: GIF, JPEG, BMP
- Otros: CAD, FIleMaker Database, Hot Potatoes, LaTeX

Por otra parte, un tema fundamental son los derechos de autor. El repositorio institucional necesita obtener permiso de los titulares para realizar acciones de difusión, reproducción y modificación/transformación. Además, este permiso posiblemente tendrá que extenderse a terceros, como los proveedores de servicios de preservación.

En cuanto a la utilización de metadatos en repositorios institucionales, una de las primeras medidas de un sistema de preservación digital es la asignación de metadatos a los objetos digitales. La clasificación más comúnmente aceptada de metadatos es la siguiente:

1. Metadatos descriptivos: representan los datos sobre el contenido intelectual que ayudan a identificar y localizar un recurso. El estándar más utilizado es Dublin Core.
2. Metadatos administrativos. son aquellos esquemas que describen la procedencia de un objeto digital, los procesos realizados para su creación o generación, sus características técnicas, sus condiciones de acceso y derechos de propiedad intelectual, y las acciones ya realizadas o previstas relacionadas con la preservación del objeto mismo.
3. Metadatos estructurales: son aquellos que se refieren a la estructura del recurso y a los elementos que lo integran.

Casos de estudio de preservación digital en repositorios nacionales e institucionales

Repositorio UNAM

Una de las funciones principales de un repositorio es contar con planes de preservación digital, es decir, que a través de metodologías y tecnologías se garantice el acceso futuro a los bienes digitales alojados en él. La Dirección General de Repositorios Universitarios (DGRU) participa en la creación de estrategias institucionales para preservar los bienes digitales de la Universidad.

La DGRU es miembro del Grupo de Preservación Digital de la Biblioteca Nacional de México, a cargo de la Universidad Nacional Autónoma de México (UNAM), donde se identifican y desarrollan diversas líneas de investigación sobre políticas y metodologías de preservación digital y se valoran distintas herramientas para enfrentar los cambios tecnológicos que podrían impedir el acceso a los recursos preservados.

Para la preservación de los objetos, el Grupo de Preservación Digital tiene como base el modelo OAIS, siendo necesario, además:

- Definir una estrategia institucional para la preservación digital.
- Definir las políticas para la preservación.
- Adecuar las políticas de digitalización.

- Adecuar las políticas de recepción de documentos.
- Adecuar las políticas de catalogación.

A continuación se muestra la arquitectura utilizada para la preservación digital:

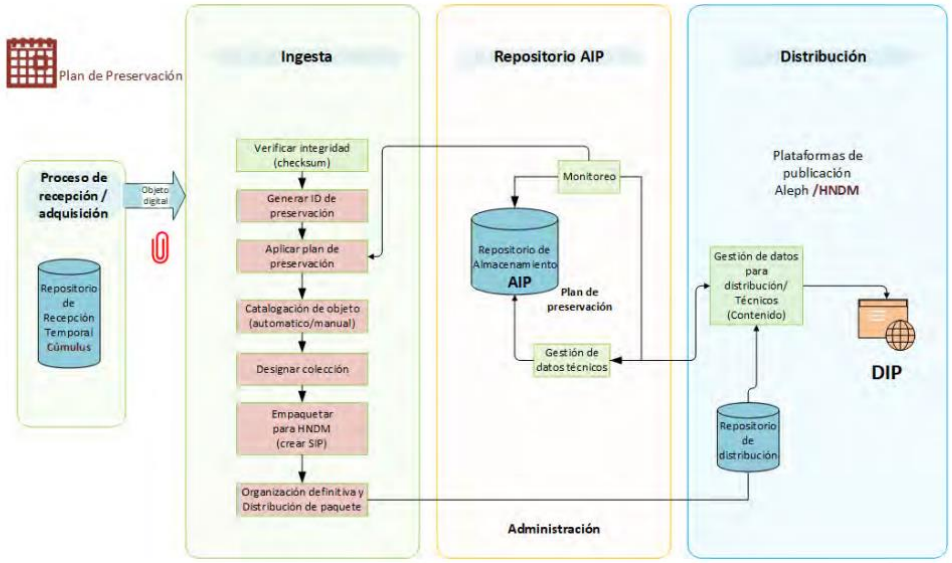

Figura 4 Arquitectura para depósito (Ramírez Molina, Ana Yuri, 2021)

Wheaton College Digital Repository

Los objetos digitales se gestionan utilizando el modelo de referencia OAIS, un marco conceptual para el archivo de material digital que también es un estándar internacional reconocido.

Preparación para la incorporación

A medida que los productores se preparan para transferir su material digital a un repositorio, lo acompañarán con los metadatos adecuados que facilitarán el acceso a largo plazo al material. Los curadores ya habrán evaluado el material por su importancia utilizando sus propios criterios.

Incorporación

Es el proceso de llevar el material digital y los metadatos correspondientes (conocido como SIP) de un productor al repositorio.

Metadatos

Los metadatos son información estructurada sobre el material y son fundamentales para preservar y proporcionar acceso a los recursos digitales y archivos de la biblioteca de Wheaton College. Los recursos digitales deben incluir metadatos de preservación

esenciales, que comprenden metadatos administrativos, metadatos técnicos, metadatos estructurales, procedencia y derechos.

Almacenamiento de archivos

Este paso se encarga del almacenamiento, el mantenimiento y la recuperación de los AIP. Una vez que se crean los AIP, se asignan al almacenamiento permanente de acuerdo con diferentes criterios (formatos, tasas de uso esperadas, etc.) El almacenamiento de archivos requiere una infraestructura técnica especializada, como la duplicación de contenido digital en sistemas tanto locales como geográficamente eliminados.

Gestión de datos

La función de gestión de datos coordina la información descriptiva asociada con los AIP de un repositorio. En particular, esta función mantiene y administra bases de datos que contienen información descriptiva y ejecuta las solicitudes de búsqueda recibidas de los usuarios. También realiza actualizaciones en las bases de datos, incluida la adición de nueva información descriptiva.

Administración

Esta función gestiona las operaciones regulares del repositorio. Esto incluye negociar acuerdos de donantes con productores, monitorear el control de acceso y ofecer servicios a los usuarios. La función desarrolla políticas y estándares, y realiza ingeniería de sistemas.

Acceso

Ayuda a los usuarios ("consumidores" en el modelo OAIS) a encontrar información relevante sobre material digital en un repositorio y acceder al material.

Estándares de preservación

El contenido digital viene en una variedad de formatos digitales e incluye tanto material digital nacido como elementos analógicos digitalizados para su preservación y acceso. Para garantizar el almacenamiento a largo plazo y el acceso al contenido digital, el repositorio adoptará estándares de formato de archivo consistentes para la preservación de copias maestras de contenido digital de acuerdo con la Iniciativa de Directrices Digitales de Agencias Federales.

MIT institutional repository

El repositorio institucional del MIT está comprometido a garantizar el acceso a largo plazo al contenido digital en cualquier formato, modelar buenas prácticas en preservación digital y alinearse con los estándares y prácticas prevalecientes para la preservación digital a medida que surgen, actualmente tiene como base el modelo OAIS, pero proporciona solamente un marco conceptual para los metadatos de preservación. Varios grupos e instituciones de todo el mundo se han basado en ese marco para identificar y especificar elementos de metadatos. Cada uno adoptó un enfoque ligeramente diferente y desarrolló diferentes especificaciones de elementos de metadatos.

El MIT trabaja con el estándar METS para la recopilación de metadatos. El esquema METS es un marco XML flexible diseñado para almacenar metadatos administrativos, estructurales y descriptivos sobre objetos digitales. Es un contenedor basado en XML para todo tipo de metadatos, para las relaciones entre ellos y los objetos sobre los que tratan, y para los comportamientos asociados con los objetos. La amplitud de METS y la flexibilidad diseñada en su estructura lo convierten en una excelente opción para un marco o contenedor para los objetos y metadatos en un sistema de preservación.

El departamento Digital Preservation Management (Gestión de la Preservación Digital) del MIT ha desarrollado varios recursos para la preservación digital. Uno de ellos es un flujo de trabajo para administrar contenido digital y un conjunto de diez principios de preservación digital:

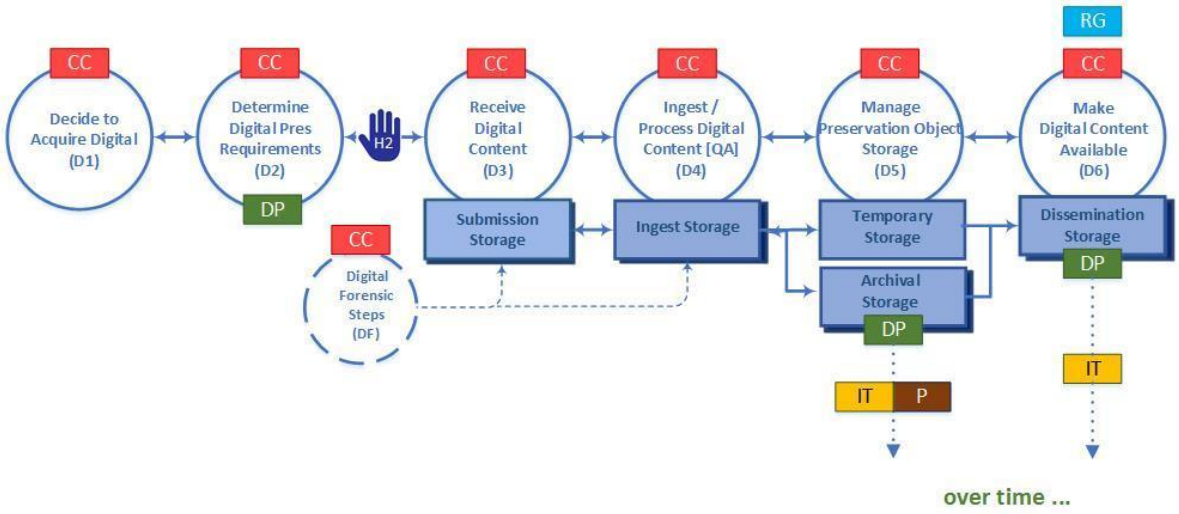

Figura 5 Flujo de trabajo para la administración de contenido digital (MIT Libraries, 2018)

V. Recomendaciones de preservación digital para el repositorio de Angola

El objetivo primero de un sistema de preservación digital es que la información contenida en él se mantenga accesible para los usuarios por un largo periodo. Esto significa que el sistema de preservación debería:

- No tener, como mínimo, un único punto de falla. Tiene que tolerar la falla de cada uno de los componentes. En general, debería ser capaz de soportar la falla de varios componentes simultáneamente.
- Ejecutar al mismo tiempo el soporte, el software y el hardware. Una vez que fallen o se vuelvan obsoletos, habrá que sustituirlos.
- Procurar hacer comprobaciones regulares en intervalos frecuentes para mantener la probabilidad de fallo en niveles aceptables.

En el sistema del diseño de preservación hay que tener en cuenta las amenazas a las que hay que hacer frente y que tienen que estar identificadas:

- Fallo de la maquinaria
- Fallo del programa
- Errores en las comunicaciones
- Fallo de los servicios de red
- Obsolescencia de los soportes
- Obsolescencia del software
- Error del operador
- Desastre natural
- Ataque externo
- Ataque interno
- Fallo organizativo

De acuerdo con la información vista a lo largo de este documento, es fundamental establecer un conjunto de buenas prácticas para aplicar un plan de preservación digital y protegerse de las amenazas, tales como:

- Escribir e implementar una política de preservación por cada colección digital.
- Identificar qué colecciones serán mantenidas a lo largo del tiempo.
- Establecer los servicios de conservación del contenido y el uso de métodos apropiados para asegurar que las condiciones se cumplen.
- Preservar la vieja tecnología que aún es funcional o que no es factible que cambie, manteniendo los equipos y programas que crearon los objetos digitales.
- Emular la vieja tecnología que es totalmente necesaria, simulando el comportamiento del software original con el que se crearon los objetos digitales.
- Migrar sistemas y formatos de datos cuando cambia la tecnología para permitir el acceso con la nueva tecnología.
- Controlar el material utilizando metadatos estructurados y otros documentos que faciliten el acceso y ayuden durante todo el proceso de preservación como lo es el uso de los metadatos en formatos XML.
- Utilizar arquitecturas de preservación, formatos de almacenamiento y metadatos estándares.
- Elegir los medios apropiados para proporcionar acceso pese a los cambios tecnológicos.
- Hacer uso de las características de DSpace para la preservación digital:
    - Uso de PREMIS y METS. DSpace ha definido un set llamado “Technical Metadata Element” para cumplir con la preservación y con las necesidades de gestión del ciclo de vida de la información.

    - Configurar los tipos de archivo según los tres niveles de conservación, admitido, conocido o no admitido.
    - Verificar la integridad de los flujos de bits en el almacén de activos con el verificador Checksum.
    - Realizar la replicación (copia de seguridad/ restauración/auditoría) del contenido de DSpace en otras ubicaciones mediante Replication Task Suite.
    - Opcionalmente, se puede utilizar JHOVE para la extracción automática de metadatos.
- En cuanto a la accesibilidad de los dispositivos de hardware, software y comunicaciones, utilizar protocolos abiertos asegura la interoperabilidad del repositorio y, por tanto, el intercambio y transferencia de información.

En el siguiente enlace http://dspace.uces.edu.ar:8180/jspui/help/formats.jsp#top se muestran los formatos de archivo soportados por Dspace. De manera resumida, y siguiendo las recomendaciones que se han expuesto anteriormente, se sugiere utilizar los siguientes formatos de archivos para los documentos del Repositorio Nacional de Angola.

- Texto: PDF
- Video: Flash, QuickTime
- Imágenes: TIFF, JPEG, PDF
- Audio: MP3

Adicionalmente, se recomienda incluir además de los formatos anteriores, el archivo XML que contenga los metadatos del documento, preferentemente estructurado con el estándar Dublin Core, para garantizar la permeabilidad del contenido.

VI. Conclusión

Los repositorios tienen una oportunidad única en el área de la preservación digital. Sería ideal lograr conservar nuestros materiales de patrimonio cultural para la perpetuidad, pero dadas las circunstancias actuales del panorama digital, es poco probable que así sea; sin embargo, podemos usar técnicas para estar más cerca de ese objetivo. Estas técnicas son muy variadas y responden a diferentes situaciones y líneas estratégicas (copias de seguridad, copia de datos en soportes durables, migración, replicación, emulación, transferencia de datos, etc.), aunque, en general, están destinadas a mantener los objetos digitales y sus características de acceso a largo plazo.

Por sí solas, las soluciones técnicas no son suficientes para asegurar la duración prolongada de los objetos digitales. Para lograr soluciones plenas y satisfactorias, se requiere la integración de aspectos técnicos y administrativos: recursos humanos, capacitación, requisitos financieros, criterios de selección, metadatos de preservación, etcétera.

Para una administración efectiva de las colecciones digitales, se debe desarrollar y seguir un plan de gestión en los proyectos, que permita evaluar los requisitos de preservación y el acceso a largo plazo. Simultáneamente, deben ser identificados los costos y los beneficios, además de estimarse los riesgos.

Para facilitar la obtención e introducción de metadatos de preservación son necesarias herramientas de extracción automática de metadatos de los archivos de origen. Estas herramientas deberían estar encadenadas a las herramientas de software que manejan los repositorios institucionales. Existen desarrollos de sistemas y herramientas asociados a diferentes componentes del modelo OAIS en forma separada e integrada que pueden ser aplicados, por ejemplo, herramientas de extracción automatizada de metadatos y de gestión de riesgos de obsolescencia de formatos.

Bibliografía

PREMIS Working Group (2008). Data Dictionary for Preservation Metadata, version 2.0. Dublin (Ohio): OCLC; Mountain View (California): RLG. http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

Digital Preservation Coalition (2008). Digital Preservation Handbook. Disponible en http://www.dpconline.org/vendor-reports/download-document/299-digital-preservation-handbook.html

MIT Libraries. (2018). Digital Preservation. https://libraries.mit.edu/about/strategic-initiatives/digital-preservation/

CCSDS Secretariat. (2002). Reference Model for an Open Archival Information System (OAIS). https://siarchives.si.edu/sites/default/files/pdfs/650x0b1.PDF

Chapter 2. DSpace System Documentation: Functional Overview. (2002). depts.washington.edu. http://depts.washington.edu/cmditr/dspace/dspace-1.5.1-release/dspace/docs/html/ch02.html

PREMIS Working Group (2008). Data Dictionary for Preservation Metadata, version 2.0. Dublin (Ohio): OCLC; Mountain View (California): RLG. http://www.loc.gov/standards/premis/v2/premis-2-0.pdf