

# Análisis documental y preservación de contenidos multimedia

Noemí Mases Blanch

PID\_00202593



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

# Índice

<b>Introducción</b> .....	5
<b>Objetivos</b> .....	6
<b>1. El análisis documental de contenidos multimedia</b> .....	7
1.1. ¿En qué consiste el análisis documental? .....	7
1.2. El etiquetado como descripción básica de los contenidos multimedia .....	9
1.3. La descripción detallada de los contenidos multimedia .....	9
1.3.1. El análisis de la imagen .....	9
1.3.2. El análisis del sonido .....	11
1.3.3. El análisis verbal .....	11
1.3.4. La interactividad .....	12
1.4. Herramientas para la descripción documental: tesauros, taxonomías, ontologías y <i>topic maps</i> .....	13
1.5. El concepto de metadatos .....	16
1.6. Estructuras de metadatos .....	18
1.7. Creación y almacenamiento de metadatos .....	19
1.8. Tipología de metadatos .....	19
<b>2. La preservación de contenidos digitales multimedia</b> .....	21
2.1. ¿Qué entendemos por preservación digital? .....	21
2.2. Los repositorios digitales fiables ( <i>trustworthy digital repositories</i> ) .....	22
2.3. Funciones de un repositorio digital fiable: el modelo OAIS .....	22
2.4. Estrategias técnicas de preservación digital .....	24
2.5. Ejemplos de software para la preservación digital .....	27
2.6. La preservación digital en otros ámbitos de actuación .....	30
2.7. Políticas de preservación digital pioneras en todo el mundo .....	31
<b>Bibliografía</b> .....	33



## Introducción

En este módulo, se explica en qué consisten dos procesos clave para la difusión, el acceso y la perdurabilidad de los contenidos multimedia a largo plazo: el análisis documental y la preservación digital.

La primera parte se centra en las dos tareas básicas que permiten este análisis, la identificación y la descripción de los contenidos, para que estos puedan ser recuperados y/o reutilizados en un futuro por los usuarios que los necesiten, sean empresas privadas, instituciones públicas o particulares. Desde el primer momento, se hace una diferenciación clara entre la descripción que realizan los usuarios que no son profesionales (a través del etiquetado social en la Red) y los profesionales de la documentación (a través de sistemas de gestión de contenidos multimedia mucho más sofisticados). En este segundo tipo de descripción, se tendrán que tener en cuenta los elementos principales que conforman un documento multimedia: la imagen, el sonido, la palabra y la interactividad, además de todos sus elementos relacionados. Por otro lado, se explican con detalle algunas de las herramientas de soporte que se utilizan para llevar a cabo esta descripción detallada de manera eficiente: los tesauros, las taxonomías, las ontologías y los *topic maps*. Por último, se analiza el concepto de metadatos, el elemento mínimo que se utiliza para identificar y describir contenidos de cualquier tipo. Presta especial atención a su estructura y tipologías, pero también a su complejo proceso de creación y almacenamiento.

La segunda parte del módulo analiza con detalle en qué consiste la preservación digital de los contenidos multimedia. En el primer apartado se define como el proceso de gestión necesario para asegurar el mantenimiento y el acceso continuo a los materiales nacidos en formato digital a largo plazo. Lo ideal sería que se llevara a cabo en todas las etapas del ciclo de vida de los documentos multimedia, es decir, desde su creación hasta su uso futuro. Es por ello que se pide la implicación de todas las instituciones, desde las empresas privadas, las administraciones públicas, etc. que los crean, hasta las bibliotecas, los archivos, los centros de documentación, las universidades, los museos, etc. que los custodian. Asimismo, se habla de “repositorios digitales fiables” para designar las bibliotecas y archivos nacionales, universidades y otros organismos académicos, si bien tienen como objetivo salvaguardar a largo plazo todos los recursos digitales que gestionan. El apartado siguiente expone las diferentes estrategias de preservación llevadas a cabo por estos repositorios y, a continuación, se explican algunos ejemplos de software que podemos encontrar hoy en día disponibles, tanto gratuitos como de pago. Por último, se analizan otros ámbitos de actuación de la preservación digital y se citan algunas de las instituciones e iniciativas pioneras a escala internacional, llegando a la conclusión de que en nuestro país todavía falta un gran camino por recorrer.

## Objetivos

Con el estudio de este módulo, lograréis los objetivos siguientes:

- 1.** Entender en qué consiste el análisis documental de los contenidos multimedia.
- 2.** Tener claro que lo conforman dos tareas básicas: la identificación y la descripción de contenidos.
- 3.** Tomar conciencia de que el etiquetado social realizado por usuarios no profesionales es también una manera de describir los contenidos multimedia.
- 4.** Comprender cómo se realiza el proceso de descripción de contenidos multimedia por parte de los profesionales de la documentación.
- 5.** Conocer las herramientas de descripción documental necesarias para llevar a cabo este proceso.
- 6.** Entender qué significa el concepto de metadatos y conocer sus estructuras, tipologías y todo su proceso de creación y almacenamiento.
- 7.** Tomar conciencia de la importancia de la preservación de los contenidos digitales multimedia para garantizar su perdurabilidad futura.
- 8.** Conocer las diferentes técnicas de preservación llevadas a cabo tanto por las empresas privadas como por las instituciones que custodian los documentos multimedia.
- 9.** Conocer algunos tipos de software de preservación digital que hay en el mercado, tanto de pago como gratuitos.
- 10.** Adquirir conciencia de los diferentes ámbitos donde se puede aplicar un sistema de preservación digital (bibliotecas y archivos nacionales, universidades, archivos administrativos, archivos personales, sector artístico, sector aeronáutico, sector científico, etc.).
- 11.** Conocer las instituciones y los proyectos líderes en preservación digital en todo el mundo.
- 12.** Saber en qué situación se encuentra nuestro país en referencia a este tema.

# 1. El análisis documental de contenidos multimedia

## 1.1. ¿En qué consiste el análisis documental?

Entendemos como análisis documental de contenidos multimedia la fase en que estos se identifican y se describen, con el objetivo de que puedan ser reutilizados (en su totalidad o en fragmentos) para su emisión, la venta o cesión a otras empresas, instituciones o particulares, su investigación, etc.

Las dos tareas básicas que permite el análisis de contenidos son:

1) **Identificación:** que permitirá el seguimiento y control de los contenidos multimedia, ya estén alojados en servicios gratuitos de intercambio de vídeos, en sistemas de gestión de contenidos/ documentos o en servicios de documentación tradicionales.

2) **Descripción:** que permitirá tratar un contenido bastante complejo, si bien puede ser interpretado y utilizado de muchas maneras. Con la irrupción de internet como fuente de información y herramienta documental, ha surgido la necesidad de crear dos tipos de descripciones muy diferenciadas:

- **Simples:** idóneas para los contenidos alojados en servicios gratuitos de intercambio (Youtube, Vimeo, etc.) y utilizados por los usuarios no profesionales. Un ejemplo de estos tipos de descripciones sería el etiquetado social.
- **Complejas:** realizadas por los profesionales de la documentación, permiten hacer análisis más detallados de los contenidos (descriptorios, resúmenes, etc.). Actualmente, y además de los tesauros, han proliferado herramientas que mejoran la descripción documental, como por ejemplo las taxonomías y las ontologías, que explicaremos más adelante.

El elemento mínimo que se utiliza para identificar y describir contenidos se denomina **metadatos**. Estos añaden información adicional sobre aspectos formales y de significado y se usan para cualquier tipo de formato (textual, oral, multimedia, icónico, etc.).

Por otro lado, hay que tener en cuenta que el análisis de contenido de los mensajes multimedia considera tres aspectos de gran importancia (Pinto Molina, García Marco y Agustín Lacruz, 2002):

1) Los **medios primarios** que conforman el mensaje: imagen, sonido no verbal y mundo no verbal. Para los documentos que transmitan sensaciones cinestésicas, gustativas u olfativas, el análisis será mucho más complejo.

2) Las **interacciones** entre los tres medios (imagen, sonido y palabra) y sus efectos posteriores. Una imagen en movimiento, por ejemplo, puede inducir efectos cinestésicos, como pasa con muchos efectos especiales de las películas.

3) Las **propiedades sistémicas** del mensaje multimedia, es decir, los marcos de comprensión, la veracidad, la intencionalidad, el tema, la relación implícita entre el mensaje, el medio, el nivel textual, el nivel conceptual, el género, el argumento, las abstracciones reales o imaginarias, los conceptos genéricos y específicos, etc.

Figura 1. Los medios que intervienen en los principales tipos de documentos analizados por las empresas audiovisuales, bibliotecas, archivos, centros de documentación, etc.

Tipo de documentos	Medios que intervienen	Ejemplo
Documentos textuales escritos	Textos + ilustraciones	Manual de comunicación no verbal
Documentos textuales orales (archivos orales)	Textos + paralingüística + sonido ambiental + [música]	Entrevista en la calle
Imagen fija	<i>Imagen fija</i> + movimientos expresados en potencia + [textos escritos]	Fotografía con pie de foto o con un cartel dentro de la imagen
Imagen móvil	<i>Imagen fija</i> (1) + ... + imagen fija (n) + <i>movimiento</i> + [textos escritos] + [música]	Película muda
Audiovisual	<i>Imagen fija</i> (1) + ... + imagen fija (n) + <i>movimiento</i> + paralingüística + sonido ambiental + [textos escritos y/o orales] + [música]	Película
Multimedia interactivo	<i>Imagen fija</i> (1) + ... + imagen fija (n) + <i>movimiento</i> + paralingüística + sonido ambiental + [textos escritos y/o orales] + [música] + manipuladores + espacio de transformaciones	Atlas interactivo

Fuente: Pinto Molina, García Marco y Agustín Lacruz (2002).

La inmediatez y la movilidad de los documentos multimedia hacen que el proceso de análisis documental sea bastante dificultoso, a diferencia de lo que pasa con los documentos textuales y de imagen fija.



## 1.2. El etiquetado como descripción básica de los contenidos multimedia

La manera más simple de describir los documentos multimedia es asignar, en lenguaje libre, una serie de términos que describan aspectos formales o de contenido. Los aspectos formales incluyen, por ejemplo, las dimensiones de reproducción, la duración y el formato del fichero que contiene el documento. Los aspectos de contenido se describirán a partir de palabras clave, sean controladas o libres, para garantizar una buena recuperación. En los servicios gratuitos de alojamiento de vídeos por internet, por ejemplo, los usuarios pueden añadir términos en lenguaje libre para describir los personajes que aparecen, la temática tratada, el equipo técnico y artístico que ha participado, etc. Estos términos se denominan etiquetas (*tags*). El **etiquetado social**, es decir, el uso de etiquetas comunes para la descripción de materiales de temática similar, es una tendencia que está muy en alza. De hecho, estos servicios de alojamiento elaboran listas básicas de términos, resultado del etiquetado de los usuarios. Estas listas se denominan **folcsonomías**, y organizan los términos alfabéticamente o a través de nubes de etiquetas o *tag clouds*, donde se destacan por medida de letra o colores especiales.

Otra manera de etiquetar que ofrecen estos servicios es por medio de las **valoraciones**. Estas valoraciones permiten crear categorías de “favoritos”, “los más vistos”, etc. Por otro lado, el hecho de compartir contenidos en redes sociales (a través de la incrustación de enlaces) como Facebook o Twitter también se considera otra manera de etiquetar, ya que los usuarios están dando a conocer sus vídeos preferidos y, por lo tanto, dan importancia a los descriptores utilizados.

### 1.3. La descripción detallada de los contenidos multimedia

Como ya hemos dicho anteriormente, la descripción de contenidos multimedia puede ser simple (realizada por usuarios no profesionales) y compleja (realizada por profesionales de la documentación). En este segundo caso, el análisis se realiza en el marco de los sistemas de gestión documental y tiene en cuenta varios elementos básicos que veremos a continuación.

#### 1.3.1. El análisis de la imagen

Para el análisis documental de las imágenes, los profesionales tienen que tener en cuenta una serie de **elementos estructurales** (Pinto Molina y otros, 2002):

a) **Ambientación**. Hace referencia al uso de la luz, el color y las transiciones entre las formas. En el momento de realizar el análisis de la imagen en cuestión, el documentalista se fijará en la presencia o ausencia de luz y sacará de

#### Enlace recomendado

Consultad la web de Delicious, un servicio de gestión de marcadores sociales (etiquetas) en la Red. Permite añadir los marcadores que tradicionalmente se guardaban en los navegadores y categorizarlos a través de folcsonomías. Además de almacenar sitios web, también permite compartirlos con otros usuarios de Delicious (<https://delicious.com>).

ello tanto significados directos (hora del día o estación del año) como interpretaciones simbólicas. El análisis de los colores también servirá para extraer significados psicológicos, emocionales y éticos.

**b) Composición.** Hace referencia a las relaciones que mantienen un conjunto de formas en un determinado espacio visual. La información que transmite puede ser perceptiva (equilibrio, movimiento), emocional (calma, tensión, debilidad, fuerza) y simbólica (relaciones personales y sociales). Además de la información, también transmite otros aspectos, como la perspectiva y el número áureo, del cual se deriva la regla de los dos tercios aplicada al cine y a la fotografía.

**c) Espacios y escenarios.** Su significado gira en torno a dos parámetros: el ser humano condicionado por su ambiente natural, y su capacidad para transformar este ambiente y crear escenarios artificiales. Por otro lado, el documentalista también tendrá que considerar el espacio de las relaciones interpersonales y toda su simbología (posición de los brazos y piernas, posición frontal o inclinada, distancia, etc.).

**d) Objetos.** Los objetos como tal son portadores de mensajes que tienen que ver tanto con la identidad cultural de las personas (según su diseño, podemos detectar la comunidad a la cual pertenecen y cuál es su lugar dentro de ella) como con sus valores individuales (a través de la decoración de los espacios, las personas emiten determinados mensajes). Otra de las propiedades que tienen los objetos es que alargan la duración del mensaje en el tiempo, a diferencia de los mensajes verbales y gestuales.

**e) Imagen humana.** El cuerpo también es un potente instrumento de comunicación. La anatomía, la fisonomía y la indumentaria nos transmitirán importantes mensajes comunicativos. Los rasgos del rostro y del cuerpo, así como las posturas, la ropa y los complementos, podrán transmitir también rasgos del carácter o del estado de ánimo de las personas.

Por otro lado, también se ha de tener en cuenta un conjunto de **elementos narrativos** (acciones) (Pinto Molina y otros, 2002):

**a) Aspectos sistémicos de la narración visual.** La narración visual intenta transmitir percepciones, sensaciones, emociones y símbolos dentro de unos límites temporales muy concretos. Por eso se recurre a procesos que suceden en la comunicación escrita y oral: descomposición de la historia a través de elipsis, planos, tomas, escenas y secuencias, y recomposición mediante elipsis, transiciones y continuidades. La ambientación y la composición ayudarán a crear sensaciones de movimiento, ritmo y tensión.

**b) Actividad expresiva (gestualidad).** Los movimientos estereotipados y los gestos y las posturas resultantes transmiten también mensajes comunicativos muy potentes. El análisis de los gestos se puede realizar del más general al

más particular. Así, se presta atención primero a los gestos del cuerpo en su conjunto, y se continúa por los brazos, las manos, el rostro y, finalmente, la mirada.

c) **Actividad instrumental y objetos.** El análisis se centra en las actividades económicas, sociales, políticas, lúdicas, etc. En el ámbito multimedia, estas actividades se usan generalmente como ambientación o pretexto. Por otro lado, las actividades instrumentales que provocan resultados inusuales, así como también el comportamiento de los objetos, requiere una especial atención por parte de los documentalistas.

### **1.3.2. El análisis del sonido**

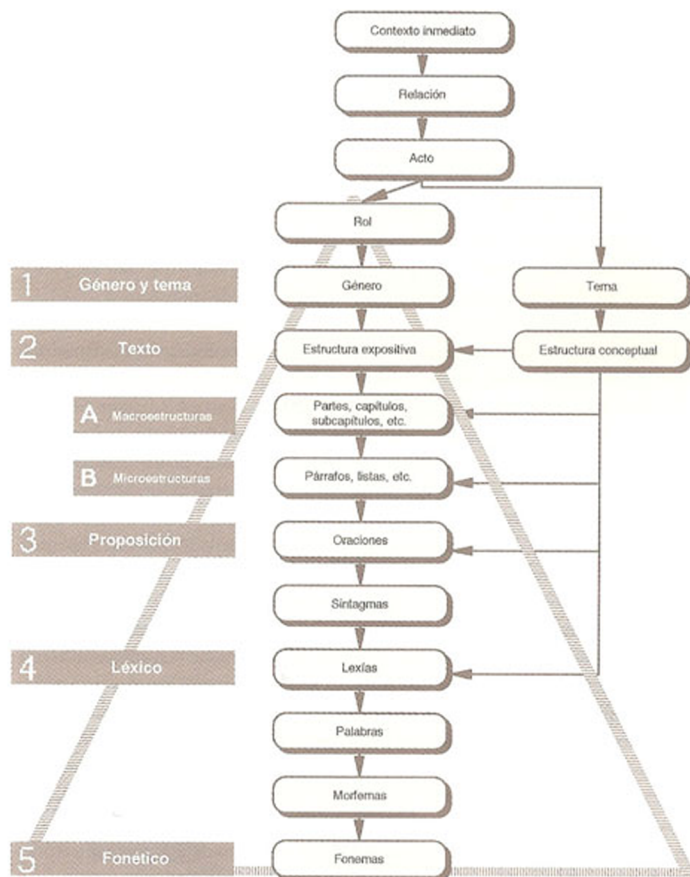
Durante el proceso de análisis del sonido se tendrán en cuenta, por un lado, el **ruido**, el **sonido** y el **silencio** y, por el otro, la **voz humana**. Es importante que durante el análisis de la banda sonora se distingan claramente los sonidos y los ruidos. Los sonidos tienen varias cualidades que transmitirán mensajes comunicativos y poderosos efectos emocionales: intensidad, duración, tono, timbre y ritmo. Por otro lado, la música es la forma no verbal más depurada de comunicación mediante sonidos. En los documentos multimedia, se utiliza por sí misma o como acompañamiento y tiene la capacidad de potenciar efectos dramáticos, creando un clima emocional, social y cultural. Por último, no tenemos que olvidar que el silencio también es un potente recurso comunicativo.

Asimismo, la voz humana como lenguaje verbal también es objeto de análisis por parte de los documentalistas, junto con sus códigos de acompañamiento, como son las características, la intensidad, el tono, el timbre, el ritmo, el arrastre, la resonancia, la colocación, la entonación, los sonidos anatómicos y emocionales (risa, llanto, suspiro, bostezo...), los símbolos orales (onomatopeyas, silbatos...), los reguladores de la conversación, etc.

### **1.3.3. El análisis verbal**

Como ya sabemos, el lenguaje verbal es un sistema complejo que funciona con un sistema de niveles organizados jerárquicamente. Los niveles de análisis se resumen en la siguiente figura:

Figura 2. Niveles de análisis del lenguaje verbal



Fuente: Pinto Molina, García Marco y Agustín Lacruz (2002).

### 1.3.4. La interactividad

La interactividad está representada en los entornos virtuales, muy importantes tanto en la educación y el ocio como en el ámbito artístico. Dentro de los procesos de análisis documental del contenido, los resúmenes tendrán que indicar las posibilidades de interacción que permiten los programas, y los descriptores el tipo de documento interactivo que se trate, representando de manera precisa estas posibilidades.

#### Lectura complementaria

Si queréis saber cómo se realiza un análisis documental detallado de fotografías, documentos audiovisuales televisivos, documentos videográficos y cinematográficos, y recursos electrónicos (CD-ROM, webs institucionales, entornos virtuales de aprendizaje, bases de datos, libros virtuales y programas de ordenador), leed los estudios de caso que propone el libro siguiente:

**M. Pinto Molina; F. J. García Marco; M. del C. Agustín Lacruz (2002).** *Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos* (pág. 217-333). Gijón: Trea.

#### **1.4. Herramientas para la descripción documental: tesauros, taxonomías, ontologías y *topic maps***

Los profesionales de la documentación utilizan varias herramientas para llevar a cabo la compleja tarea de descripción de los contenidos multimedia. Estas herramientas son las siguientes:

##### **a) Los tesauros**

A menudo se comparan con los diccionarios, pero a pesar de que contienen unidades léxicas, tienen diferencias muy claras. Los diccionarios contienen varias definiciones de una unidad léxica en un contexto lingüístico, mientras que los tesauros contienen términos de indexación y recuperación de un área del conocimiento. Nos muestran, por lo tanto, el campo de aplicación de un término y su posición en la disciplina. Por otro lado, podemos definirlos como tres tipos de lenguaje (Naumis Peña, 2007):

- Lenguaje documental, porque recogen palabras simples y compuestas que integran las expresiones utilizadas por los autores de los documentos de un área temática concreta. De este modo, actúan como términos de recuperación para quien busque información sobre contenidos temáticos.
- Lenguaje de indexación, porque permiten representar los temas de los documentos y los mensajes implícitos de estos. Por lo tanto, aseguran el tratamiento del contenido temático del documento.
- Lenguaje controlado, porque incluyen términos validados por un proceso en el que se establecen relaciones entre ellos para mantener la univocidad de sus descriptores y se determina la preferencia de un término por encima de los otros posibles.

La clasificación de la información en un tesoro se realiza a través de sistemas alfabéticos, numéricos, alfanuméricos, facetados o temáticos que sirven para agrupar los documentos en un orden clasificatorio después de someterlos al proceso de indexación. Del mismo modo que los vocabularios controlados o las listas de encabezamientos, los tesauros normalizan los temas que contienen los documentos. A diferencia de la indexación por palabras clave, la información indexada a partir de términos normalizados es mucho más precisa de cara a la recuperación, ya que informa de los términos que no están aceptados y, por lo tanto, no tienen que ser utilizados, y también de todas las temáticas que puede contener el documento sin que hayan sido expresadas de manera implícita en el contenido de este.

Figura 3. Resultado de la busca realizada bajo el término *artista* en el tesoro de la Unesco


Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

## Thesaurus de la UNESCO

3 entradas correspondientes a : **artista** Nueva búsqueda (ayuda)

Para obtener las noticias indizadas con un término dado en unesdoc/unesbib, haga click en el [número]

Término : **Artista** [549] ⊖

English term: Artists  
 Terme français: Artiste  
 Русский термин: Художники

MT 3.45 Arte  
 UP Diseñador gráfico  
 UP Fotógrafo  
 UP Grafista  
 TE Arquitecto [95]  
 ....TE2 Arquitecto restaurador [11]  
 TE Artista escénico [138]  
 ....TE2 Actor [67]  
 ....TE2 Bailarín [26]  
 ....TE2 Músico [192]  
 .....TE3 Compositor [82]  
 TE Director de cine [129]  
 TE Director de teatro [16]  
 TE Escultor [36]  
 TE Mujer artista [64]  
 TE Pintor [116]  
 TR Personal cultural [27]  
 TR Profesión [224]

Término : **Artista del espectáculo** ⊖

Fuente: <http://databases.unesco.org/thessp>

### Enlace recomendado

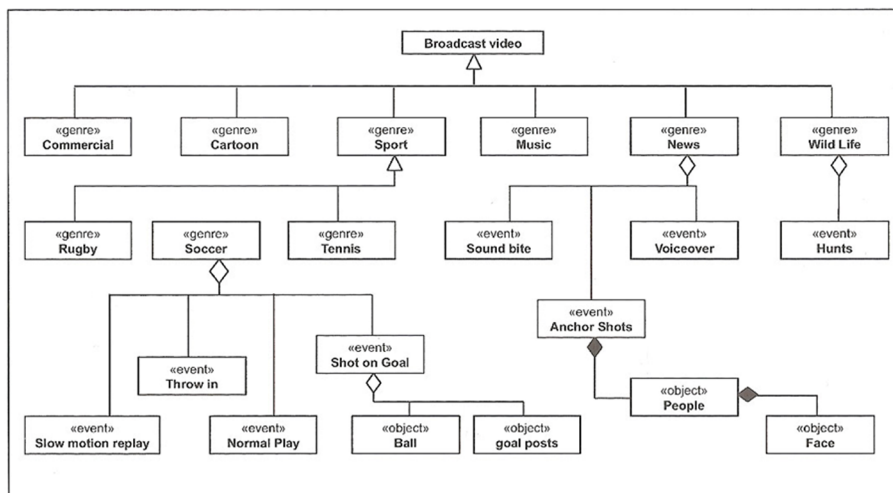
Si queréis conocer más a fondo la estructura y el funcionamiento de un tesoro, consultad el tesoro de la Unesco, que contiene miles de términos controlados relacionados con el ámbito de la educación, la cultura, las ciencias naturales, las ciencias sociales y humanas, la comunicación y la información: <http://databases.unesco.org/thessp>

### b) Las taxonomías

Se trata de un conjunto de términos organizados de manera jerárquica y a la vez clasificados por categorías y subcategorías, de forma que cada término está relacionado con los términos de los que depende y con los términos que dependen de él. Así pues, las taxonomías presentan relaciones semánticas de los términos de manera más amplia que los tesauros, pues pueden incluir datos corporativos en los términos y en la propia megaestructura donde se añaden, además, varios tesauros. En definitiva, pueden abrazar toda la información sobre una organización/institución, es decir, sus procesos de negocio, la normativa y pautas sobre la manera de trabajar del personal, los contenidos que trata, etc.

Normalmente, las empresas del sector audiovisual acostumbran a utilizar para su trabajo interno un conjunto de clasificaciones jerárquicas aplicadas, por ejemplo, a los tipos de programas que producen, a los personajes que forman parte de ellas o a los tipos de formatos de transmisión técnica de los contenidos. Desgraciadamente, no suelen relacionar cada una de estas clasificaciones entre sí para formar una taxonomía única, ya que es una tarea bastante costosa. Lo cierto es que, de momento, no existe una taxonomía muy definida y aceptada para los objetos digitales actuales. Es un tema que los expertos están debatiendo a fondo.

Figura 4. Extracto de una taxonomía



Fuente: Roach y otros (2002).

### Enlaces recomendados

Si queréis conocer más a fondo el funcionamiento de una taxonomía, consultad las News-Codes, creadas por el International Press Telecommunications Council (IPTC). Las agencias de prensa y los medios de comunicación las usan para la asignación de metadatos sobre los tipos de géneros periodísticos, temas, planos de imágenes fijas o vídeo, etc.: <http://www.iptc.org/site/newscodes/overview>

Leed este interesante blog donde se explican, desde un punto de vista particular, aspectos relacionados con el uso de las taxonomías en las organizaciones, en las redes sociales, etc.: <http://thetaxonomyblog.com>

Consultad la web del Taxonomy Boot Camp de Washington, un congreso anual donde se muestran los últimos avances en organización, gestión e integración de información y datos en las empresas. Va dirigido especialmente a los profesionales de la información, como por ejemplo gestores de contenidos, especialistas en taxonomías y clasificaciones, arquitectos de la información, ingenieros del conocimiento, creadores de portales e intranets, etc.: <http://www.taxonomybootcamp.com>

### c) Las ontologías

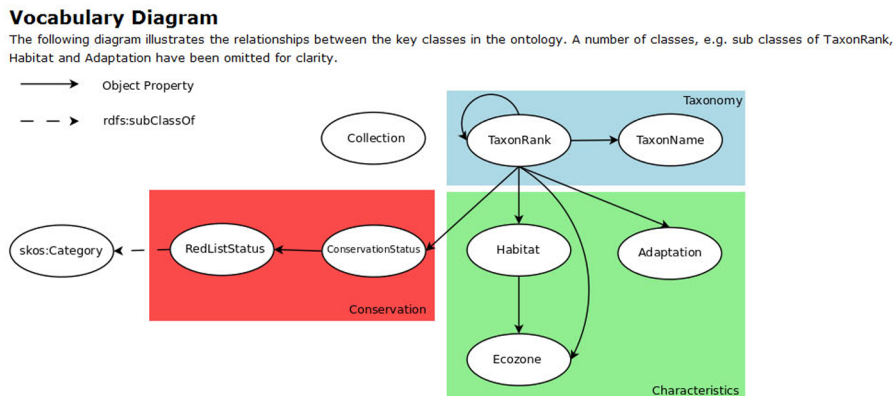
Las ontologías definen los posibles tipos de contenidos existentes en un fondo documental digital, es decir, sus características y elementos, así como también las diferentes relaciones que se establecen entre estos. No todas estas relaciones tienen que ser aplicadas a todos los elementos y características, del mismo modo que un mismo elemento o característica puede tener más de una relación.

Estas herramientas de descripción documental son bastante complejas de diseñar, pues definen una gran cantidad y variedad de relaciones. Es por eso por lo que son utilizadas, en la mayoría de los casos, por grandes instituciones. Ejemplos de ello serían la BBC y la European Broadcasting Union (EBU).

En la Red, el concepto de ontología va estrechamente ligado a la Web semántica, que pretende convertir los recursos que se encuentran disponibles en la Red en una gran fuente de información, accesible a través de todo tipo de herramientas automatizadas. Podríamos decir que las ontologías cumplen funciones de diccionarios para etiquetar el lenguaje web con precisión, establecer sus propiedades y sus relaciones con otros conceptos y transformarlos en

términos, para que las herramientas de busca interpreten el significado de las palabras encontradas. Son, por lo tanto, diferentes a los motores de busca, que funcionan con palabras clave.

Figura 5. Diagrama que ilustra las relaciones entre las clases clave en la Wildlife Ontology, creada por la BBC



### Overview Of Terms

An alphabetical index of the ontology terms, divided into classes, properties and individuals. All the terms are hyperlinked to their detailed description for quick reference.

**Classes:** | [Adaptation](#) | [ExtremesAdaptation](#) | [AnimallIntelligence](#) | [BehaviouralPattern](#) | [Class](#) | [Collection](#) | [CommunicationAdaptation](#) | [ConservationStatus](#) | [EcosystemRole](#) | [Ecozone](#) | [Family](#) | [FeedingHabit](#) | [FreshwaterHabitat](#) | [Genus](#) | [Habitat](#) | [Kingdom](#) | [LifeCycle](#) | [LocomotionAdaptation](#) | [MarineHabitat](#) | [MorphologyAdaptation](#) | [Order](#) | [Phylum](#) | [PredationStrategy](#) | [RedListStatus](#) | [ReproductionStrategy](#) | [SocialBehaviour](#) | [Species](#) | [SurvivalStrategy](#) | [TaxonName](#) | [TaxonRank](#) | [TerrestrialHabitat](#)

Fuente: <http://www.bbc.co.uk/ontologies/wildlife/2010-02-22.shtml>

## d) Los *topic maps* (mapas temáticos)

Se trata de un estándar que se está implantando en la web y que posibilita la navegación conceptual. Están relacionados con la Web semántica y se desarrollan en lenguajes como XML y XTM. Los términos que se extraen durante el proceso de indexación de los documentos tienen relación con un término genérico y términos asociados, igual que los tesauros. La ubicación en el ámbito de conocimiento estará definida por las relaciones que tienen con los términos que se encuentran en el propio documento. Ahora bien, estos términos no serán unívocos, como pasa con los tesauros, sino que se usarán todos los sinónimos que aparezcan en el documento y variarán en función del lenguaje utilizado por cada autor. En definitiva, los *topic maps* no son vocabularios definidos previamente, sino que son un modelo flexible con el vocabulario del propio documento indexado. Por otro lado, tanto los tesauros como las taxonomías y las ontologías pueden ser una fuente muy valiosa para la construcción de los *topic maps* (Naumis Peña, 2007).

### 1.5. El concepto de metadatos

Como ya hemos comentado al inicio de este apartado, los metadatos son el elemento mínimo que se utiliza para identificar y describir contenidos digitales. Se trata de un conjunto de datos que aportan información sobre cada documento. En el caso de un documento multimedia, sería el conjunto de datos que identifican y describen la combinación de vídeo, audio, texto, gráficos, y que giran en torno a las características físicas, la duración de todo el conjunto, el lugar de realización, las descripciones del título o el tema, etc. Se utilizan en

### Enlace recomendado

Para saber cómo funciona una ontología, consulta la web de la Wildlife Ontology, desarrollada por la BBC: <http://www.bbc.co.uk/ontologies/wildlife/2010-02-22.shtml>

Para entender las diferencias entre el funcionamiento de un tesoro y una ontología, lee el artículo siguiente:

L. Codina; R. Pedraza-Jiménez (2011, septiembre-octubre). "Tesauros y ontologías en sistemas de información documental". *El profesional de la información* (vol. 20, núm. 5, pág. 555-563). [http://www.lluiscodina.com/ontologiatesauros\\_2011.pdf](http://www.lluiscodina.com/ontologiatesauros_2011.pdf)

### Lectura complementaria

Si queréis saber más sobre los *topic maps*, lee el artículo:

R. Eíto Brun (2003, marzo-abril). "Topic maps y la indexación de recursos electrónicos en la web". *El profesional de la información* (vol. 12, núm. 2, pág. 141-148). <http://www.elprofesionalde la informacion.com/contenidos/2003/marzo/15.pdf>



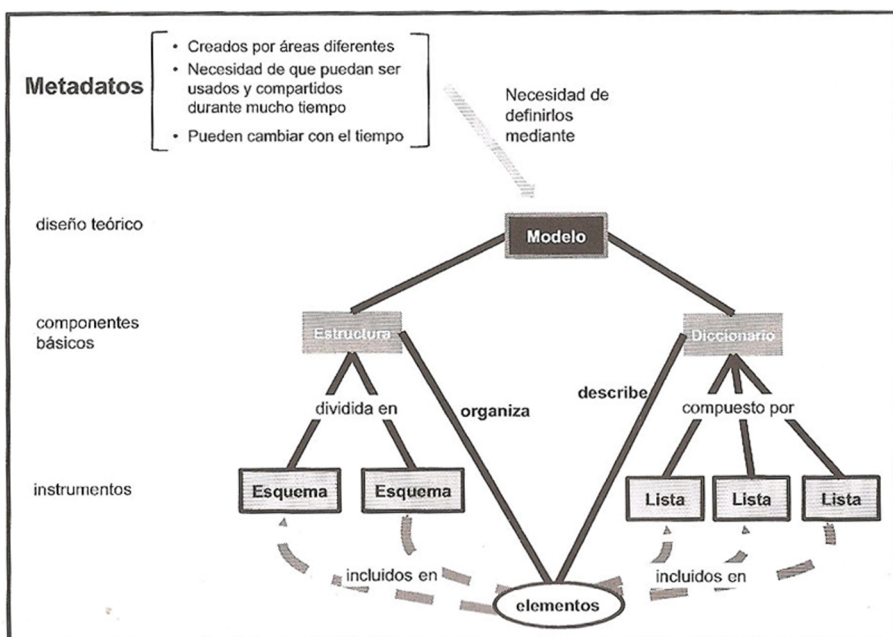
cualquier momento del flujo de trabajo, es decir, se pueden incorporar cada vez que se cree, se modifique, se relacione, se emita, se archive o se recupere un contenido.

Los metadatos tienen que cumplir unos aspectos mínimos (Caridad y otros, 2011):

- Cada una de las áreas de trabajo en una producción multimedia querrá usar aquellos metadatos que les sean más útiles. Los profesionales (ingenieros) encargados de transmitir los contenidos tendrán que tener algunos requisitos para empaquetarlos y facilitar así su difusión.
- A menudo pasa que estos contenidos se tienen que utilizar y compartir durante mucho tiempo. Esto implica, por lo tanto, que los metadatos tendrán que ser diseñados para que se mantengan en un futuro, no solamente durante el uso concreto que se pueda hacer en un determinado momento. Además, deberán ser útiles para múltiples usuarios.
- Tanto los contenidos como los metadatos ya añadidos tendrán que permitir que otros usuarios/documentalistas puedan ir añadiendo nuevos metadatos para describir el tema y los personajes con más profundidad, etc.
- En un futuro, estos metadatos se tendrán que mejorar o cambiar.

Así pues, el valor de los metadatos radica en su relación con otros metadatos, es decir, cómo se tienen que definir de forma que no se dupliquen en las diferentes áreas de trabajo, en qué orden tienen que ser introducidos, cuál tiene que ser su grado de complejidad, si tienen que contener lenguaje controlado o libre, etc. En definitiva, los metadatos que se utilicen tendrían que tener los componentes siguientes:

Figura 6. Componentes básicos de un modelo de metadatos



Fuente: M. Caridad y otros, 2011.

Lo cierto es que todo el conjunto de metadatos de un sistema multimedia se tendría que recoger en un modelo teórico que especificara todos los requisitos necesarios para su uso. Por un lado, tendrían que organizarse en una estructura común, pues a menudo son utilizados por varios usuarios. Por el otro, todos los elementos que conforman los metadatos tendrían que estar recogidos en un diccionario que permitiera identificar el nombre asignado a cada elemento de manera clara, definir por qué se utiliza este elemento y unificar criterios para evitar que un nombre concreto se utilice para elementos diferentes.

### 1.6. Estructuras de metadatos

Cualquier estructura de metadatos tendría que cumplir los siguientes requisitos (Rust y Bide, 2000):

**a) Identificación única:** cada contenido tiene que incluir al menos un identificador, que a la vez tiene que cumplir una serie de condiciones:

- Ha de ser estable, es decir, no puede ser modificado. Ahora bien, puede ser un identificador relativo. Esto significa que si un objeto cambia físicamente de ubicación (por ejemplo, dentro de una base de datos accesible a través de internet), los usuarios pueden utilizar el mismo identificador para acceder al contenido, mediante un sistema de redirecciones.
- Tiene que ser único dentro de un sistema de identificación. Por ejemplo, si un contenido está formado por audio y vídeo, se podrá utilizar un identificador para todo el conjunto, otro para el audio, y otro para el vídeo.
- Tiene que ser seguro, es decir, se deben utilizar algoritmos y técnicas de encriptado para que los programas que usen los identificadores realicen comprobaciones digitales y los den por válidos.
- Tiene que proporcionar acceso a metadatos básicos que sirvan para describir el contenido identificado.

**b) Asignación de autoridad:** esto significa que cada vez que se introduzca o se modifique un metadato, se tendrá que identificar claramente al autor.

**c) Granularidad funcional:** en las producciones multimedia es muy útil identificar tanto el contenido en su conjunto como cada uno de sus posibles fragmentos. Esto se realizará desde diferentes puntos de vista: técnico (secuencias, planos), semántico (cortes separados por tema o personaje), temporal (detectar el comienzo de cada inicio de fragmento dentro de un programa determinado, por ejemplo), entre otros.

**d) Permiso de acceso:** es importante dar acceso a los metadatos que necesiten los usuarios, pero también se tiene que restringir a todo el personal que no esté autorizado, bien porque no tenga los conocimientos técnicos suficientes, bien porque sean datos de entrada restringidos.

## 1.7. Creación y almacenamiento de metadatos

Durante el proceso de creación multimedia, es preferible que la inserción de los metadatos interfiera lo menos posible. Actualmente, muchas herramientas de filmación y grabación ya incluyen la posibilidad de añadir automáticamente metadatos relacionados con la fecha de la toma, la duración, el código de tiempo de cada imagen captada, la localización geográfica, etc. Por otro lado, muchos sistemas de gestión de contenidos también los generan automáticamente (por ejemplo, reconociendo elementos que aparecen repetidamente en diferentes imágenes, el grado de luminosidad que permite distinguir la localización y el momento del día, etc.). Por último, hay maneras de añadir metadatos de manera semiautomática, escogiendo los términos de entre una serie cerrada de opciones definida previamente a través de un vocabulario controlado.

El almacenamiento de estos metadatos también es un proceso importante a tener en cuenta. Se lleva a cabo de dos maneras:

- Guardando en un mismo soporte físico todo el contenido, es decir, tanto el material como los metadatos. Serían los llamados metadatos incrustados o *embedded*. Es la opción más adecuada y es muy útil, sobre todo cuando los metadatos no cambien en el tiempo.
- Guardando por separado el material y los metadatos, manteniendo su relación. En este caso hablamos de metadatos externos. Este sistema es útil para metadatos descriptivos o que se tengan que modificar a menudo.

Sea como fuere, lo importante es mantener el código que une cada esencia (es decir, cualquier dato necesario para representar una experiencia sensorial, visual, oral o de cualquier otro tipo) con sus respectivos metadatos. Si esto no funciona, se hará difícil localizar y reconocer el contenido que se está buscando. Este código se establecerá a partir de dos maneras (Caridad y otros, 2011):

- Mediante XML, el lenguaje de marcado de texto que constituye una serie de reglas básicas que facilitan la definición de estructuras de metadatos y los metadatos basados en estas estructuras.
- Mediante el formato *KLV* (*key, length, value*), enfocado principalmente para la transmisión de contenido. Cada metadato se encuentra dentro de un campo numérico, del cual se describen tres características: la clave (*key*), que identifica el contenido; la longitud (*length*) numérica del campo; y el valor (*value*) que se da para este campo.

## 1.8. Tipología de metadatos

Existen varios tipos de metadatos:

a) **Metadatos físicos**, que proporcionan información sobre la ubicación de los ficheros, las formas de almacenamiento, datos sobre autenticidad y seguridad, los formatos de digitalización, las ratios de compresión, la velocidad de reproducción, los tipos de decodificadores de reproducción, la resolución, la frecuencia de muestreo, el formato de salida, la cantidad de pistas de audio, etc.

b) **Metadatos administrativos**, utilizados para la gestión y administración de los contenidos digitales, proporcionan información sobre el método de adquisición, los derechos, los criterios de selección del contenido, el control de las diferentes versiones, los usuarios internos que han accedido a ellos, etc. También ofrecen información sobre las condiciones de uso, si se trata de un original o una copia, qué características materiales e informáticas tiene (soportes para su reproducción, por ejemplo), etc.

c) **Metadatos descriptivos**, utilizados tradicionalmente por los documentalistas, informan sobre el autor, el título, la duración, la fecha de producción, el lugar de edición, la ubicación física donde se realizó la grabación y/o la grabación, etc. También se añaden los descriptores temáticos, tanto palabras clave en lenguaje libre como descriptores extraídos de lenguajes documentales controlados (tesauros, listas jerarquizadas de términos, taxonomías, etc.), así como también transcripciones de audio, entre otros datos.

## 2. La preservación de contenidos digitales multimedia

### 2.1. ¿Qué entendemos por preservación digital?

Como sabemos, la información en papel y otros soportes, como por ejemplo los microfilmes, pueden durar centenares y miles de años, en algunos casos. En cambio, la información codificada en formato digital es bastante improbable que dure más de una década o dos. Las razones principales son diversas: el envejecimiento de los soportes digitales, la obsolescencia de los formatos, del software y del hardware, y los problemas de incompatibilidad de los nuevos sistemas (Bia y Sánchez, 2002).

Definimos la *preservación digital* como la gestión necesaria para asegurar el mantenimiento y el acceso continuo a los materiales creados en formato digital (conocidos también como *born digital*) a largo plazo.

Para garantizar este acceso, habrá que tener en cuenta lo siguiente (Keefer y Gallart, 2005):

- La evolución de la tecnología.
- La coordinación de todas las instituciones implicadas (*stakeholders*) durante el proceso de creación, distribución, almacenamiento y uso de estos contenidos. Estas instituciones van desde las empresas y las administraciones públicas hasta las asociaciones profesionales, las bibliotecas, los archivos, etc.
- El desarrollo y la aplicación de normativas adecuadas para llevar a cabo el proceso.

Cuando hablamos de preservación digital no nos referimos al proceso de digitalización de contenidos ni a realizar copias de seguridad. Estas últimas se llevan a cabo, básicamente, para proteger los documentos ante pérdidas de datos por culpa de apagones o por roturas de soportes. Lo que se guarda es tanto la información publicada en el servidor de la institución o empresa como los recursos digitales en proceso de edición. La preservación digital, en cambio, se ocupa de salvaguardar los recursos digitales que necesitaremos en un futuro.

Un ejemplo que permite ver esta diferencia es el caso de las imágenes gráficas. Por un lado, interesa realizar copias de seguridad de las que están en formato JPG de baja calidad guardadas en el servidor, y por el otro, interesa realizar

<sup>(1)</sup>Las técnicas empleadas, en este segundo caso, las veremos en el apartado 2.4.

copias de preservación de las que están en formato TIFF de alta calidad que no se encuentran en el servidor (por su peso y por la lentitud de transmisión), pero que seguramente se necesitarán para usos futuros. En definitiva, existen diferencias notables entre los dos procesos. Si bien las copias de seguridad se realizan diariamente o semanalmente, las copias de preservación se realizarán normalmente cada año o año y medio<sup>1</sup> (Bia y Sánchez, 2002)

## **2.2. Los repositorios digitales fiables (*trustworthy digital repositories*)**

Los repositorios digitales fiables son instituciones que tienen como objetivo preservar a largo plazo los recursos digitales que custodian. Estas instituciones van desde bibliotecas y archivos nacionales hasta universidades y otros organismos académicos.

Los repositorios digitales fiables serán responsables de lo siguiente:

- Mantener la infraestructura técnica adecuada.
- Garantizar la integridad y la autenticidad de los recursos.
- Vigilar la protección de los derechos intelectuales.
- Limitar su acceso solo a los usuarios autorizados.
- Mantener las herramientas de identificación y recuperación de la información.
- Suministrar los metadatos normalizados y actualizados que tendrán que acompañar la información.

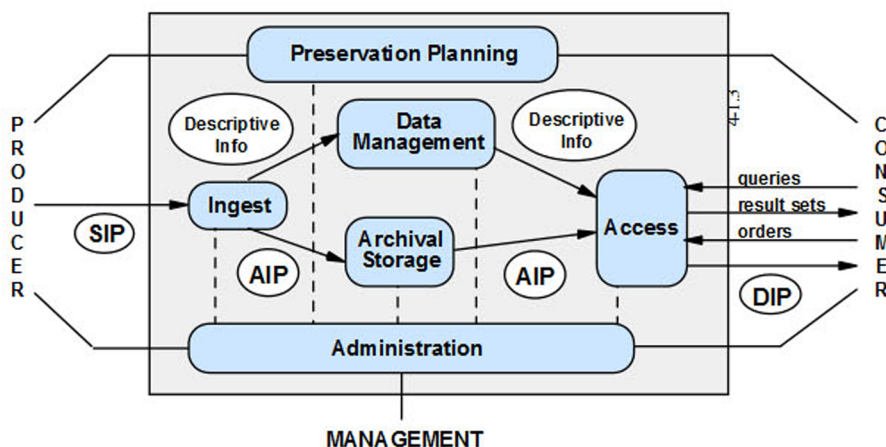
Existen iniciativas que evalúan estos repositorios. Se trata de auditorías y certificaciones que ofrecen sistemas para medir la adecuación y precisión de los procesos que llevan a cabo (planes de formación, currículums del personal, descripciones de trabajo, etc.). El objetivo es analizar la infraestructura organizativa, la gestión de los contenidos digitales, la tecnología, la infraestructura técnica y la seguridad que ofrecen estas instituciones. Algunas de estas certificaciones son TRAC (*Trustworthy Repositories Audit & Certification*, <http://www.crl.edu/pdf/rac.pdf>), NESTOR (*Catalog of Criteria for Trusted Digital Repositories*, <http://edoc.hu-berlin.de/series/nestor-materialien/8en/pdf/8en.pdf>) y DRAMBORA (*Digital Repository Audit Method Based on Risk Assessment*, <http://www.repositoryaudit.eu>).

## **2.3. Funciones de un repositorio digital fiable: el modelo OAIS**

En el año 2002, la NASA publicó el **Open Archival Information System (OAIS)** (<http://ddp.nist.gov/refs/oais.pdf>), un modelo aceptado y estandarizado como Norma ISO 14721, que establece las funciones que tiene que cumplir un sistema de preservación digital, así como también la manera de gestionar sus ficheros y metadatos. Estas funciones son básicamente seis:

- Ingesta: entrada de datos y ficheros. Es uno de los módulos más desarrollados de los sistemas de preservación. A través de políticas coherentes, los productores multimedia tendrían que crear ficheros en formatos estándares.
- Almacenamiento: grabación de datos.
- Gestión de datos: introducción de los metadatos correspondientes.
- Acceso: módulos de consulta.
- Preservación: llevada a cabo por la dirección interna de la institución encargada de la preservación. Tendrá mucha importancia la vigilancia tecnológica.
- Servicios comunes: comprobación interna, para que todo funcione correctamente.

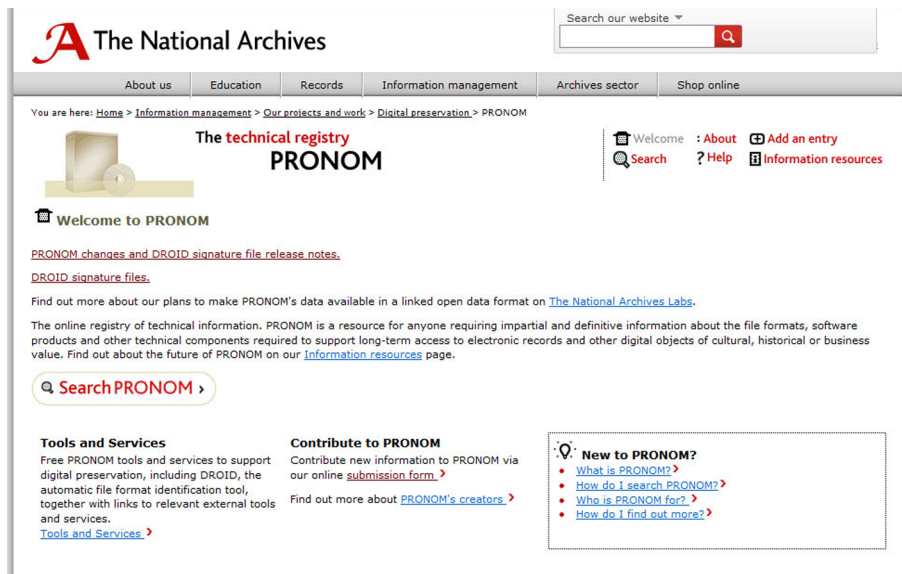
Figura 7. Diagrama donde se muestran las seis funciones del modelo OAIS, creado por la NASA



Fuente: <http://ddp.nist.gov/refs/oais.pdf>

Una de las funciones que puede presentar más problemas es la ingesta de datos. A veces, los ficheros que se tienen que introducir en un sistema de preservación digital tienen extensiones ilegibles o que se hacen difíciles de reconocer. La solución pasa por utilizar programas de detección de formatos que contienen, además, el registro de todos los formatos existentes a escala mundial. Un ejemplo es **PRONOM** (<http://www.nationalarchives.gov.uk/pronom/default.aspx>), desarrollado por The National Archives del Reino Unido.

Figura 8. Portada de PRONOM.



Otros programas que identifican y validan formatos son los siguientes:

- **DROID** (*Digital Record Object Identification*, <http://sourceforge.net/projects/droid>): gratuito y también creado por The National Archives del Reino Unido.
- **JHOVE** (*JSTOR/Harvard Object Validation Environment*, <http://jhove.sourceforge.net>): orientado a sistemas más profesionales.
- **XENA** (<http://xena.sourceforge.net>): gratuito, desarrollado por The National Archives of Australia y basado en Java.

## 2.4. Estrategias técnicas de preservación digital

Los principales implicados en la preservación de los fondos documentales son, tradicionalmente, las bibliotecas y los archivos. Ahora bien, los productores, en este caso multimedia, también pueden facilitar esta tarea desde el momento en que diseñan y crean los contenidos, poniendo especial énfasis en los formatos, la consistencia, la identificación y descripción, y la normalización. Por lo tanto, es necesario tener en cuenta cuestiones de conservación en todas las etapas del ciclo de vida de los documentos.

En este sentido, las empresas e instituciones tendrían que diseñar un plan de preservación que respondiera a las preguntas siguientes (Bia y Sánchez, 2002):

- ¿Qué se tiene que conservar y por qué?
- ¿Dónde conservarlo?
- ¿Hasta cuándo conservarlo?
- ¿Cómo encontrarlo después?
- ¿Cómo conseguir que se mantenga inalterado?
- ¿Cómo evitar que se vuelva obsoleto?

### Enlaces recomendados

Algunos de los proyectos de referencia que se basan en la elaboración y aplicación del modelo OAIS son **PLANETS** (<http://www.planets-project.eu>), y **CASPAR** (**C**ultural, **A**rtistic and **S**cientific **K**nowledge for **P**reservation, **A**ccess and **R**etrieval, <http://www.casparpreserves.eu>), ambos cofundados por la Unión Europea y otras instituciones dedicadas a la preservación digital.



Es importante que este plan sea revisado de manera periódica y asignar a los contenidos preservados un límite de vida. Por otro lado, los implicados en este proceso han de ser conscientes de que ninguno de los soportes digitales actuales garantiza una vida larga de la información que contienen. Esta información está constantemente expuesta a una serie de factores que hacen peligrar su perdurabilidad: negligencias humanas, errores técnicos y mecánicos, virus, obsolescencia, incompatibilidad y pérdida de los programas, cambios no documentados ni autorizados, metadatos incompletos, falta de actualización de la información, etc.

En el momento de llevar a cabo una política de preservación digital, las empresas e instituciones deberán analizar a fondo su viabilidad. Esto significa que tendrán que ver si son capaces de asumir y cumplir las responsabilidades asociadas a esta política. Desgraciadamente, podemos afirmar que todavía hay una carencia generalizada de financiación, de asignación de responsabilidades, de políticas muy diseñadas y de planes de formación para el personal. Por otro lado, también está toda la cuestión legal, todavía poco resuelta, y que gira en torno a lo siguiente:

- El derecho de modificación de los contenidos, es decir, el paso de un formato a otro.
- El derecho de recoger y archivar los contenidos.
- El derecho de proporcionar el acceso al público.

Una vez se haya analizado la viabilidad, la empresa o institución procederá a las diferentes estrategias de preservación que más convengan. Algunas de estas son:

#### a) *Refreshing* o actualización de los datos en un soporte nuevo

Supone una solución al problema del deterioro de los soportes digitales. Se produce a partir de la transferencia de los datos de un soporte a otro, manteniendo intacto su *bitstream* original y sin afectar al software necesario para su lectura.

#### b) Migración o cambio de formato

Es una medida contra la obsolescencia y consiste en convertir la información en nuevos formatos. A diferencia del *refreshing*, esta técnica sí que altera el *bitstream* original e implica también un cambio de software para su lectura. Uno de los riesgos más importantes es la posible pérdida de datos y de la funcionalidad del entorno original. Además, se tiene que tener en cuenta que los formatos quedan obsoletos con el paso del tiempo.

#### c) Emulación

#### Enlace recomendado

El UK Data Archive trabaja con múltiples formatos de datos para diferentes usos. Consultad este enlace que contiene los formatos de preservación aceptados por este organismo: <http://data-archive.ac.uk/create-manage/format/formats-table>.

Consiste en recrear el entorno de software y hardware. Permite, por lo tanto, que el software original sea utilizado sin necesidad de que el sistema original que lo ejecutaba siga existiendo. Es importante preservar tanto el emulador como el sistema operativo, la aplicación y los datos. Algunas arquitecturas emuladas son: Atari2600, Nintendo 64, Game Boy, PlayStation y Sega Genesis.

Como ejemplos de emuladores de videoconsolas, tenemos **MAME** (<http://www.mamedev.org>) (emulador de máquinas recreativas Arcade), **Cyberste-lla** (<http://stella.sourceforge.net>) (emulador de la videoconsola Atari 2600) y **zSNES** (<http://sourceforge.net/projects/zsnes>) (emulador de la videoconsola Super Nintendo). Otro buen ejemplo de emulador es **Dioscuri** (<http://dioscuri.sourceforge.net>), desarrollado por la Koninklijke Bibliotheek y el Nationaal Archief de los Países Bajos.

#### d) Estandarización

Se trata de buscar una manera de codificar y formatear los contenidos digitales adhiriéndose a estándares reconocidos y más utilizados. Se presupone que estos estándares perdurarán en el tiempo e irán solucionando los problemas de compatibilidad que vayan apareciendo con la evolución del entorno informático.

#### e) Encapsulación

Consiste en empaquetar la información que se tiene que preservar junto con sus metadatos descriptivos. A través de estructuras físicas o lógicas, llamadas *containers* o *wrappers*, se proporciona una relación entre todos los componentes de la información que hay que preservar. Esta estrategia pretende evitar la obsolescencia de la información digital, disminuyendo la posibilidad de que se pierdan los elementos necesarios para codificar y traducir los datos, puesto que los detalles sobre cómo interpretar los bits en cualquier sistema futuro quedan registrados.

Para la encapsulación, se ha creado un formato específico denominado **UPF** (*Universal Preservation Format*). Su objetivo es que los objetos digitales sean independientes del software y de los sistemas operativos, pero que los metadatos que traigan incorporados incluyan las especificaciones técnicas para tener acceso a los materiales encapsulados.

#### f) Ordenador virtual universal (UVC)

Se trata de la representación virtual de un ordenador simplificado que permite guardar información actual codificada para descifrarla en un futuro. Para poder funcionar, utiliza elementos tanto de la migración como de la encapsulación, a pesar de que para muchos se considera una técnica de emulación. Consiste en que un programa independiente registra, desde un principio, la

arquitectura básica de cada ordenador para poder simular sus memorias, las secuencias de registros y la manera de conectarse entre ellos, para reconstruir los ficheros en su aspecto original a través de un decodificador de formatos.

### g) Arqueología digital

Consiste en recuperar información a partir de fuentes de datos estropeados, fragmentados o arcaicos. El proceso consiste en descifrar el contenido, crear estándares de migración y copiar en formato legible para los futuros usuarios. La British Library es una de las pioneras en utilizar esta técnica de preservación.

## 2.5. Ejemplos de software para la preservación digital

La mayoría de los grandes sistemas que se encuentran hoy en día en el mercado y que son utilizados por algunas instituciones dedicadas a la preservación digital, como por ejemplo las bibliotecas o los archivos nacionales, contemplan funciones importantes como la ingesta de datos, la importación y transformación de metadatos, la migración de formatos y la implantación de políticas de vigilancia tecnológica. Sin embargo, la mayoría de estos sistemas tienen un coste económico bastante elevado. Ahora bien, también se está desarrollando software de código abierto, utilizado tanto por empresas privadas como por instituciones académicas. Veamos algunos ejemplos:

### a) Software comercial

#### Digital Archiving Information System (DAIS, <http://dais.cah.ucf.edu>)

Desarrollado por IBM y la Koninklijke Bibliotheek de los Países Bajos.

Figura 9. Portada de DAIS



#### Safety Deposit Box (SDB, <http://www.safedepositbox.com.my>)

Desarrollado por Tessella, The National Archives del Reino Unido y el National Archives and Record Administration (NARA) de Estados Unidos.

Figura 10. Portada de SDB

Rosetta (<http://www.exlibrisgroup.com/category/rosettaoverview>)

Desarrollado por Exlibris y The National Library of New Zealand.

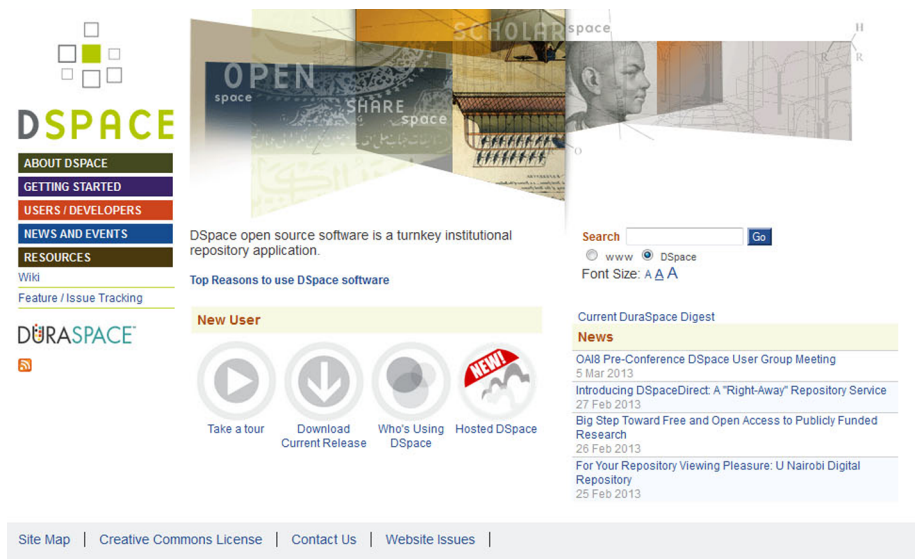
Figura 11. Portada de Rosetta

b) Software de código abierto

DSpace (<http://www.dspace.org>)

Completamente gratuito, es utilizado tanto por empresas comerciales como por instituciones académicas. DuraSpace es la organización encargada de desarrollarlo y mejorarlo.

Figura 12. Portada de DSpace



**Flexible Extensible Digital Object Repository Architecture (FEDORA, <http://fedora-commons.org>)**

También de acceso abierto, este repositorio es utilizado, sobre todo, para preservar archivos digitales personales. DuraSpace es también la organización encargada de desarrollarlo y mejorarlo.

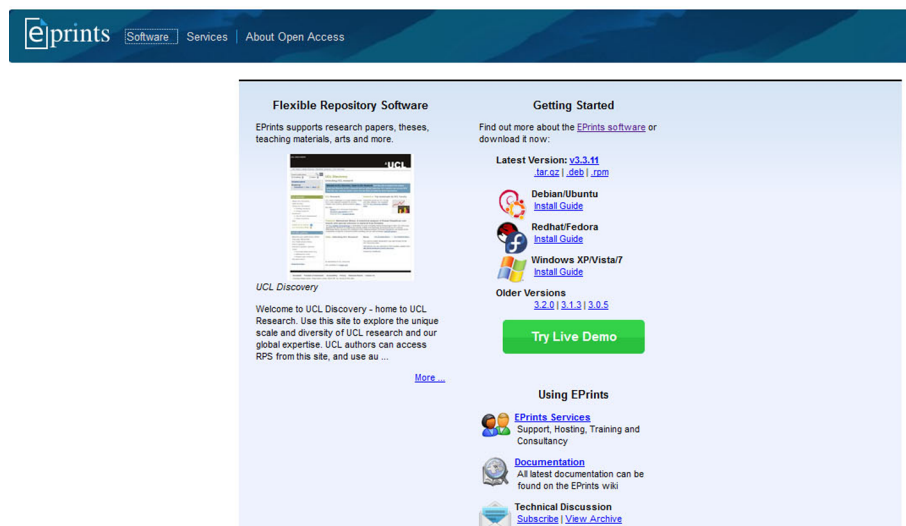
Figura 13. Portada de FEDORA



**ePrints (<http://www.eprints.org>)**

También gratuito, garantiza la preservación digital de todo tipo de documentos, desde datos científicos, tesis, informes y material para la investigación hasta multimedia.

Figura 14. Portada de ePrints



## 2.6. La preservación digital en otros ámbitos de actuación

Además de los grandes proyectos de preservación digital orientados a grandes bibliotecas y archivos, se están llevando a cabo investigaciones que se ocupan de los repositorios *open access* y de la preservación ligada a los archivos administrativos, tanto de las empresas privadas como públicas. En estas empresas, no solamente es necesario preservar la documentación que generan, sino también las firmas electrónicas y las marcas de tiempo (*data stamp*). Algunos sistemas que garantizan esta preservación son el catalán **iArxiu** (<http://www.aoc.cat/inici/serveis/gestio-interna/iarxiu>) y el alemán **Archisafe** (<http://www.archisafe.de/s/archisafe/index>) (Térmens, 2009).

En cuanto al sector artístico, uno de los comités del International Council of Museums (ICOM), el International Committee for Museum Documentation (Cidoc), está coordinando la investigación para encontrar soluciones de preservación de los museos, sobre todo en cuanto al *net.art* y la perdurabilidad de los sistemas virtuales, como pueden ser las reconstrucciones virtuales de monumentos históricos. Otra iniciativa que investiga la preservación relacionada con los museos y el arte multimedia es el proyecto **Documentation and Conservation of Media Arts Heritage (DOCAM)**, coordinado por el Centre for Research and Documentation (CR+D) de la Fondation Daniel Langlois.

Otros ámbitos en que se está investigando son la aeronáutica, para consolidar formatos, procesos y sistemas que aseguren la perdurabilidad de toda la documentación técnica de los aviones; la ciencia, donde se están elaborando importantes bases de datos sobre astronomía, meteorología, biología, genética, etc.; y los archivos personales, como por ejemplo los de los partidos políticos.

### Enlace recomendado

Si queréis saber más sobre el proyecto DOCAM, escuchad el audio de la conferencia que llevó a cabo Alain Depocas, director del CR+D de la Fondation Daniel Langlois, el 31 de mayo del 2010 en el Auditorio del Museo de Arte Contemporáneo de Barcelona (MACBA): <http://www.macba.cat/ca/documentacio-i-conservacio-de-les-arts-multimedia-centre-de-reerca-i-documentacio-crd-de-la-fundacio-daniel-langlois>

## 2.7. Políticas de preservación digital pioneras en todo el mundo

Hasta ahora, algunas bibliotecas y archivos nacionales, así como también universidades, empresas de software e instituciones depositarias de documentación, han liderado la investigación en preservación digital. Estados Unidos, Australia, Nueva Zelanda y países europeos como Países Bajos, Reino Unido, Francia y Alemania han experimentado los avances más importantes. En España, en cambio, no ha habido demasiadas iniciativas por lo que respecta a este asunto. No es hasta hace muy poco que se han empezado a llevar a cabo algunos proyectos competitivos a escala internacional.

Veamos a continuación algunas instituciones e iniciativas pioneras (Térmens, 2009):

### a) Estados Unidos

- Minerva (Library of Congress)
- National Digital Information Infrastructure and Preservation Program (Library of Congress)
- National Archives and Record Administration (NARA). Electronic Records Archives (ERA)
- LOCKKS (Stanford University)
- iRODS (University of North Carolina at Chapel Hill y University of California)
- DAITSS (Florida Center for Library Automation)
- Jhove (Harvard University)
- Unified Digital Formats Registry (Harvard University)
- Internet Archive
- Portico

### b) Australia y Nueva Zelanda

- International Internet Preservation Consortium (National Library of Australia)
- XENA (National Archives of Australia)
- PROV Digital Archive
- National Library of New Zealand

### c) Países Bajos

- E-Depot y Dioscuri (Koninklijke Bibliotheek)
- Nationaal Archief

### d) Reino Unido

- Digital Lives (British Library)
- Join Information Systems Committee (JISC)

- The National Archives
- Digital Curation Centre (CCC) de Edimburgo

#### e) Francia

- Système de Préservation et d'Archivage Réparti (SPAR) (Bibliothèque Nationale de France)
- Archives Nationales de France

#### f) Alemania

- Kopal (Deutsche Nationalbibliothek y Niedersaechsische Staats-und Universitatetsbibliothek Goettingen)
- Archisafe (Physikalisch-Technische Bundesanstalt)

#### g) España

- Biblioteca Nacional de España (forma parte del grupo de trabajo europeo Long Term Archiving and Retrieval)
- Grupo de Trabajo de Patrimonio Digital del Ministerio de Cultura
- Red de Bibliotecas Universitarias Españolas (grupo de trabajo sobre preservación digital en las bibliotecas universitarias)
- Patrimoni Digital de Catalunya (PADICAT) (Biblioteca de Catalunya)
- iArxiu (Grup de Treball d'Innovació Tecnològica de la Subdirecció General d'Arxius de la Generalitat de Catalunya)

#### **Lectura complementaria**

Si queréis conocer más sobre el proyecto Patrimoni Digital de Catalunya (PADICAT), dedicado a la preservación de la web y equiparable a las mejores iniciativas internacionales, leed el artículo siguiente:

C. Lluca (2006, noviembre-diciembre). "Archivando la Web, el proyecto Padicat (Patrimonio Digital de Cataluña)". *El profesional de la información* (vol. 15, núm. 6, pág. 473-478). <http://diposit.ub.edu/dspace/bitstream/2445/23442/1/548441.pdf>

En definitiva, la investigación en preservación digital todavía es un campo emergente y poco definido. En la práctica, todavía hay un grado de implementación bastante bajo, en el sentido de que solo algunas grandes instituciones y empresas privadas han empezado a aplicar sistemas de preservación digital, unos sistemas que todavía no se están comercializando de manera amplia. La buena noticia es que cada vez hay más conocimiento sobre este tema, y esto hace que vayan proliferando bastantes sistemas de código abierto basados en XML y Java, compatibles con la mayoría de entornos de trabajo. Por otro lado, no podemos obviar que en nuestro país existe un cierto retraso en este ámbito, y las iniciativas por parte de bibliotecas, archivos, universidades y museos son todavía insuficientes. No obstante, se espera que en un futuro no muy lejano la situación cambie, sobre todo en cuanto a las bibliotecas nacionales y universitarias, con recursos digitales cada vez más voluminosos (Térmens, 2009).



## Bibliografía

### Bibliografía básica

**Bía, A.; Sánchez, M.** (2002). *Desarrollo de una política de preservación digital: tecnología, planificación y perseverancia*. Alicante: Biblioteca Virtual Miguel de Cervantes. <<http://www.biblioteca.org.ar/libros/142322.pdf>>

**Caridad, M.; Hernández, T.; Rodríguez, D.; Pérez, B.** (2011). *Documentación audiovisual: nuevas tendencias en el entorno digital*. Madrid: Síntesis.

**Keefer, A.; Gallart, N.** (2005). "El problema de la producción y la preservación de información digital". *Comunicación audiovisual digital: nuevos medios, nuevos usos, nuevas formas*. Barcelona: UOC.

**Naumis Peña, C.** (2007). *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. México D.F.: Universidad Autónoma de México; Buenos Aires: Alfagrama.

**Pinto Molina, M.; García Marco, F. J.; Agustín Lacruz, M. C.** (2002). *Indización y resumen de documentos digitales y multimedia: técnicas y procedimientos*. Gijón: Trea.

**Térmens, M.** (2009, noviembre-diciembre). "Investigación y desarrollo en preservación digital: un balance internacional". *El profesional de la información* (vol. 18, núm. 6). <<http://eprints.rclis.org/16218/1/Investigacion-preservacion-EPI.pdf>>

**Térmens, M.; Keefer, A.** (2010). *La preservació dels recursos digitals a les biblioteques*. Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya.

### Bibliografía complementaria

**Codina, L.; Pedraza-Jiménez, R.** (2011, septiembre-octubre). "Tesauros y ontologías en sistemas de información documental". *El profesional de la información* (vol. 20, núm. 5, pág. 555-563). <[http://www.lluiscodina.com/ontologiaTesauros\\_2011.pdf](http://www.lluiscodina.com/ontologiaTesauros_2011.pdf)>

**Depocas, A.** (2010). *Documentació i conservació de les arts multimèdia: Centre de Recerca i Documentació (CR+D) de la Fondation Daniel Langlois*. Barcelona: MACBA. <<http://www.macba.cat/ca/documentacio-i-conservacio-de-les-arts-multimedia-centre-de-recerca-i-documentacio-crd-de-la-fundacio-daniel-langlois>>

**Eíto Brun, R.** (2003, marzo-abril). "Topic maps y la indización de recursos electrónicos en la web". *El profesional de la información* (vol. 12, núm. 2, pág. 141-148). <<http://www.elprofesionaldelainformacion.com/contenidos/2003/marzo/15.pdf>>

**Lluca, C.** (2006, noviembre-diciembre). "Archivando la Web, el proyecto Padicat (Patrimonio Digital de Cataluña)". *El profesional de la información* (vol. 15, núm. 6, pág. 473-478). <<http://diposit.ub.edu/dspace/bitstream/2445/23442/1/548441.pdf>>

**Roach, M. y otros** (2002). "Recent trends in video analysis: a taxonomy of video classification problems". *Proceedings of the International Conference on Internet and Multimedia Systems and Applications, IASTED*. Cambridge: IASTED. <<http://galilee.swan.ac.uk/publications/M.J.Roach/IMSA2002.pdf>>

**Rust, G.; Bide, M.** (2000). "The <indecs> metadata framework. Principles, model and data dictionary". <[http://www.doi.org/topics/indecs/indecs\\_framework\\_2000.pdf](http://www.doi.org/topics/indecs/indecs_framework_2000.pdf)>

### Enlaces

ArchiSafe: <http://www.archisafe.de/s/archisafe/index>

Archives Nationales de France: [www.archivesnationales.culture.gouv.fr/chan/chan/index.html](http://www.archivesnationales.culture.gouv.fr/chan/chan/index.html)

Biblioteca Nacional de España: <http://www.bne.es>

Catalogue of Criteria for Trusted Digital Repositories (NESTOR): <http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf>

Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval (CASPAR): <http://www.casparpreserves.eu>

Cyberstella: <http://stella.sourceforge.net>

DAITSS: <http://daitss.fcla.edu>

Delicious: <https://delicious.com>

Digital Archiving Information System (DAIS): <http://dais.cah.ucf.edu>

Digital Curation Centre (CCC): <http://www.dcc.ac.uk>

Digital Lives: <http://britishlibrary.typepad.co.uk/files/digital-lives-synthesis02-1.pdf>

Digital Record Object Identification (DROID): <http://sourceforge.net/projects/droid>

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA): <http://www.repositoryaudit.eu>

Dioscuri: <http://dioscuri.sourceforge.net>

Documentation and Conservation of Media Arts Heritage (DOCAM): <http://www.docam.ca/en.html>

DSpace: <http://www.dspace.org>

ePrints: <http://www.eprints.org>

Flexible Extensible Digital Object Repository Architecture (Fedora): <http://fedora-commons.org>

Grupo de Trabajo de Patrimonio Digital del Ministerio de Cultura: [http://www.mcu.es/bibliotecas/MC/ConsejoCB/GruposTrabajo/GT\\_Patrimonio.html](http://www.mcu.es/bibliotecas/MC/ConsejoCB/GruposTrabajo/GT_Patrimonio.html)

iArxiu: <http://www.aoc.cat/Inici/SERVEIS/Gestio-interna/iArxiu>

International Internet Preservation Consortium (IIPC): <http://netpreserve.org>

Internet Archive: <http://archive.org>

iRODS: <https://www.irods.org>

Join Information Systems Committee (JISC): <http://www.jisc.ac.uk>

JSTOR/Harvard Object Validation Environment (JHOVE): <http://jhove.sourceforge.net>

Koninklijke Bibliotheek: <http://www.kb.nl>

Kopal: <http://kopal.langzeitarchivierung.de>

LOCKKS: <http://www.lockss.org>

Long Term Archiving and Retrieval (LOTAR): <http://www.lotar-international.org>

Minerva: <http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>

Multiple Arcade Machine Emulator (MAME): <http://www.mamedev.org>

The National Archives: <http://www.nationalarchives.gov.uk>

National Archives and Record Administration (NARA). Electronic Records Archives (ERA): <http://www.archives.gov/era>

National Archives of Australia: <http://www.naa.gov.au>

National Digital Information Infrastructure and Preservation Program: <http://www.digitalpreservation.gov>

National Library of Australia: <http://www.nla.gov.au>

National Library of New Zealand: <http://natlib.govt.nz>

Nationaal Archief: <http://www.nationaalarchief.nl>

NewsCodes (International Press Telecommunications Council): [www.iptc.org/site/NewsCodes/Overview](http://www.iptc.org/site/NewsCodes/Overview)

Open Archival Information System: <http://ddp.nist.gov/refs/oais.pdf>

Patrimoni Digital de Catalunya (PADICAT): <http://www.padicat.cat/ca>

PLANETS: <http://www.planets-project.eu>

Portico: <http://www.portico.org/digital-preservation>

PRONOM: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

PROV Digital Archive: <http://prov.vic.gov.au/government/vers/history/digital-archive>

Red de Bibliotecas Universitarias Españolas (REBIUN): <http://www.rebiun.org/Repositorios>

Rosetta: <http://www.exlibrisgroup.com/category/RosettaOverview>

Safety Deposit Box (SDB): <http://www.safedepositbox.com.my>

Système de Préservation et d'Archivage Réparti (SPAR): [http://www.bnf.fr/fr/professionnels/conserver\\_spar.html](http://www.bnf.fr/fr/professionnels/conserver_spar.html)

The Taxonomy Blog: <http://thetaxonomyblog.com>

Taxonomy Boot Camp: <http://www.taxonomybootcamp.com>

Tesaurus de la Unesco: <http://databases.unesco.org/thessp>

Trustworthy Repositories Audit & Certification (TRAC): <http://www.crl.edu/pdf/rac.pdf>

UK Data Archive: <http://data-archive.ac.uk>

Unified Digital Formats Registry: <http://www.udfr.org>

Wildlife Ontology (BBC): <http://www.bbc.co.uk/ontologies/wildlife/2010-02-22.shtml>

XENA: <http://xena.sourceforge.net>

ZSNES: <http://sourceforge.net/projects/zsnes>

