

Anàlisi documental i preservació de continguts multimèdia

Noemí Mases Blanch

PID_00202575



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

Introducció	5
Objectius	6
1. L'anàlisi documental de continguts multimèdia	7
1.1. En què consisteix l'anàlisi documental?	7
1.2. L'etiquetatge com a descripció bàsica dels continguts multimèdia	9
1.3. La descripció detallada dels continguts multimèdia	9
1.3.1. L'anàlisi de la imatge	9
1.3.2. L'anàlisi del so	11
1.3.3. L'anàlisi verbal	11
1.3.4. La interactivitat	12
1.4. Eines per a la descripció documental: tesaurus, taxonomies, ontologies i <i>topic maps</i>	12
1.5. El concepte de metadades	17
1.6. Estructures de metadades	18
1.7. Creació i emmagatzematge de metadades	19
1.8. Tipologia de metadades	20
2. La preservació de continguts digitals multimèdia	21
2.1. Què entenem per preservació digital?	21
2.2. Els repositoris digitals fiables (<i>trustworthy digital repositories</i>)	22
2.3. Funcions d'un repositori digital fiable: el model OAIS	22
2.4. Estratègies tècniques de preservació digital	24
2.5. Exemples de programari per a la preservació digital	26
2.6. La preservació digital en altres àmbits d'actuació	29
2.7. Polítiques de preservació digital pioneres arreu del món	30
Bibliografia	33

Introducció

En aquest mòdul s'explica en què consisteixen dos processos clau per a la difusió, l'accés i perdurabilitat dels continguts multimèdia a llarg termini: l'anàlisi documental i la preservació digital.

La primera part se centra en les dues tasques bàsiques que permeten aquesta anàlisi, la identificació i la descripció dels continguts, per tal que aquests puguin ser recuperats o reutilitzats en un futur pels usuaris que els necessitin, siguin empreses privades, institucions públiques o particulars. Val a dir que, des del primer moment, es fa una diferenciació clara entre la descripció que fan els usuaris que no són professionals (per mitjà de l'etiquetatge social a la Xarxa) i els professionals de la documentació (per mitjà de sistemes de gestió de continguts multimèdia molt més sofisticats). En aquest segon tipus de descripció s'hauran de tenir en compte els elements principals que conformen un document multimèdia: la imatge, el so, la paraula i la interactivitat, a més de tots els seus elements relacionats. D'altra banda, s'expliquen amb detall algunes de les eines de suport que s'utilitzen per a dur a terme aquesta descripció detallada de manera eficient: els tesaurus, les taxonomies, les ontologies i els *topic maps*. Finalment, s'analitza el concepte de metadades, l'element mínim que s'utilitza per a identificar i descriure continguts de qualsevol tipus. Es fa especial atenció en la seva estructura i en les tipologies, però també en el seu complex procés de creació i emmagatzematge.

La segona part del mòdul analitza amb detall en què consisteix la preservació digital dels continguts multimèdia. En el primer apartat es defineix el procés de gestió necessari per a assegurar el manteniment i l'accés continu als materials nascuts en format digital a llarg termini. L'ideal seria que es dugués a terme en totes les etapes del cicle de vida dels documents multimèdia, és a dir, des de la creació fins a l'ús futur. És per això que es demana la implicació de totes les institucions, des de les empreses privades, les administracions públiques, etc., que els creen, fins a les biblioteques, els arxius, els centres de documentació, les universitats, els museus, etc., que els custodien. D'altra banda, es parla de "repositoris digitals fiables" per a designar les biblioteques i arxius nacionals, universitats i altres organismes acadèmics, per bé que tenen com a objectiu salvaguardar a llarg termini tots els recursos digitals que gestionen. L'apartat següent exposa les diferents estratègies de preservació dutes a terme per aquests repositoris i, a continuació, s'expliquen alguns exemples de programari que podem trobar avui dia disponibles, tant gratuïts com de pagament. Finalment, s'analitzen altres àmbits d'actuació de la preservació digital i s'esmenten algunes de les institucions i iniciatives pioneres a escala internacional, tot arribant a la conclusió que al nostre país encara manca un gran camí per recórrer.

Objectius

Amb l'estudi d'aquest mòdul, assolireu els objectius següents:

- 1.** Entendre en què consisteix l'anàlisi documental dels continguts multimèdia.
- 2.** Tenir clar que el conformen dues tasques bàsiques: la identificació i la descripció de continguts.
- 3.** Prendre consciència que l'etiquetatge social fet per usuaris no professionals és també una manera de descriure els continguts multimèdia.
- 4.** Comprendre com fan el procés de descripció de continguts multimèdia els professionals de la documentació.
- 5.** Conèixer les eines de descripció documental necessàries per a dur a terme aquest procés.
- 6.** Entendre què significa el concepte de metadades i conèixer-ne les estructures, tipologies i tot el procés de creació i emmagatzematge.
- 7.** Prendre consciència de la importància de la preservació dels continguts digitals multimèdia per tal de garantir-ne la perdurabilitat futura.
- 8.** Conèixer les diferents tècniques de preservació dutes a terme tant per les empreses privades com per les institucions que custodien els documents multimèdia.
- 9.** Conèixer alguns tipus de programari de preservació digital que hi ha al mercat, tant de pagament com gratuïts.
- 10.** Adquirir consciència dels diferents àmbits on es pot aplicar un sistema de preservació digital (biblioteques i arxius nacionals, universitats, arxius administratius, arxius personals, sector artístic, sector aeronàutic, sector científic, etc.).
- 11.** Conèixer les institucions i els projectes capdavanters en preservació digital arreu del món.
- 12.** Saber en quina situació es troba el nostre país en referència a aquest tema.

1. L'anàlisi documental de continguts multimèdia

1.1. En què consisteix l'anàlisi documental?

Entenem com a anàlisi documental de continguts multimèdia la fase en què aquests s'identifiquen i es descriuen, amb l'objectiu que puguin ser reutilitzats (totalment o en fragments) per a l'emissió, la venda o cessió a altres empreses, institucions o particulars, la investigació, etc.

Les dues tasques bàsiques que permeten l'anàlisi de continguts són:

1) **Identificació:** que permetrà el seguiment i control dels continguts multimèdia, ja estiguin allotjats en serveis gratuïts d'intercanvi de vídeos, en sistemes de gestió de continguts/ documents o en serveis de documentació tradicionals.

2) **Descripció:** que permetrà tractar un contingut força complex, per bé que pot ser interpretat i utilitzat de moltes maneres. Amb la irrupció d'Internet com a font d'informació i eina documental, ha sorgit la necessitat de crear dos tipus de descripcions ben diferenciades:

- **Simple:** idònies per als continguts allotjats en serveis gratuïts d'intercanvi (Youtube, Vimeo, etc.) i utilitzats pels usuaris no professionals. Un exemple d'aquests tipus de descripcions seria l'etiquetatge social.
- **Complex:** fetes pels professionals de la documentació, permeten fer anàlisis més detallades dels continguts (descriptors, resums, etc.). Actualment, i a banda dels tesaurus, han proliferat eines que milloren la descripció documental, com ara les taxonomies i les ontologies, que explicarem més endavant.

L'element mínim que s'utilitza per a identificar i descriure continguts s'anomena **metadada**. Aquesta afegeix informació addicional sobre aspectes formals i de significat i es fa servir per a qualsevol tipus de format (textual, oral, multimèdia, icònic, etc.).

D'altra banda, cal tenir en compte que l'anàlisi de contingut dels missatges multimèdia considera tres aspectes de gran importància (Pinto Molina i altres, 2002):

1) Els **mitjans primaris** que conformen el missatge: imatge, so no verbal i món no verbal. Per als documents que transmetin sensacions cinestèsiques, gustatives o olfactivas, l'anàlisi serà molt més complexa.

2) Les **interaccions** entre els tres medis (imatge, so i paraula) i els seus efectes posteriors. Una imatge en moviment, per exemple, pot induir efectes cinestèsics, com passa amb molts efectes especials de les pel·lícules.

3) Les **propietats sistèmiques** del missatge multimèdia, és a dir, els marcs de comprensió, la veracitat, la intencionalitat, el tema, la relació implícita entre el missatge, el medi, el nivell textual, el nivell conceptual, el gènere, l'argument, les abstraccions reals o imaginàries, els conceptes genèrics i específics, etc.

Figura 1. Els mitjans que intervien en els principals tipus de documents analitzats per les empreses audiovisuals, biblioteques, arxius, centres de documentació, etc.

Tipo de documentos	Medios que intervienen	Ejemplo
Documentos textuales escritos	Textos + ilustraciones	Manual de comunicació 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

Font: Pinto Molina i altres (2002)

La immediatesa i la mobilitat dels documents multimèdia fa que el procés d'anàlisi documental sigui força difícil, a diferència del que passa amb els documents textuais i d'imatge fixa.

1.2. L'etiquetatge com a descripció bàsica dels continguts multimèdia

La manera més simple de descriure els documents multimèdia és assignar, en llenguatge lliure, una sèrie de termes que descriguin aspectes formals o de contingut. Els aspectes formals inclouen, per exemple, les dimensions de reproducció, la durada i el format del fitxer que conté el document. Els aspectes de contingut es descriuran a partir de paraules clau, siguin controlades o lliures, per tal de garantir-ne una bona recuperació. En els serveis gratuïts d'allotjament de vídeos per Internet, per exemple, els usuaris poden afegir termes en llenguatge lliure per descriure els personatges que hi apareixen, la temàtica tractada, l'equip tècnic i artístic que hi ha participat, etc. Aquests termes s'anomenen *etiquetes* (*tags*). L'**etiquetatge social**, és a dir, l'ús d'etiquetes comunes per a la descripció de materials de temàtica similar, és una tendència que està molt a l'alça. De fet, aquests serveis d'allotjament elaboren llistes bàsiques de termes, resultat de l'etiquetatge dels usuaris. Aquestes llistes s'anomenen **folksonomies**, i organitzen els termes alfabèticament o per mitjà de *núvols d'etiquetes* o *tag clouds*, en què es destaquen per mida de lletra o colors especials.

Una altra manera d'etiquetar que ofereixen aquests serveis és per mitjà de les **valoracions**. Aquestes valoracions permeten crear categories de "favorits", "els més vistos", etc. D'altra banda, el fet de compartir continguts en xarxes socials (per mitjà de la incrustació d'enllaços) com Facebook o Twitter també es considera una altra manera d'etiquetar, ja que els usuaris estan donant a conèixer els seus vídeos preferits i, per tant, donen importància als descriptors utilitzats.

1.3. La descripció detallada dels continguts multimèdia

Com ja hem comentat anteriorment, la descripció de continguts multimèdia pot ser simple (feta per usuaris no professionals) i complexa (feta per professionals de la documentació). En aquest segon cas, l'anàlisi es fa en el marc dels sistemes de gestió documental i té en compte diversos elements bàsics que veurem a continuació.

1.3.1. L'anàlisi de la imatge

Per a l'anàlisi documental de les imatges, els professionals han de tenir en compte una sèrie d'**elements estructurals** (Pinto Molina i altres, 2002):

a) **Ambientació**. Fa referència a l'ús de la llum, el color i les transicions entre les formes. En el moment de fer l'anàlisi de la imatge en qüestió, el documentalista es fixarà en la presència o absència de llum i n'extraurà tant significats directes (hora del dia o estació de l'any) com interpretacions simbòliques. L'anàlisi dels colors també servirà per a extreure'n significats psicològics, emocionals i ètics.

Enllaç recomanat

Consulteu el web de Delicious, un servei de gestió de marcadors socials (etiquetes) a la Xarxa. Permet afegir els marcadors que tradicionalment es desaven en els navegadors i categoritzar-los per mitjà de folksonomies. A més d'emmagatzemar llocs web, també els permet compartir amb altres usuaris de Delicious (<https://delicious.com>).

b) Composició. Fa referència a les relacions que mantenen un conjunt de formes en un determinat espai visual. La informació que transmet pot ser perceptiva (equilibri, moviment), emocional (calma, tensió, debilitat, força) i simbòlica (relacions personals i socials). A més de la informació, també transmet altres aspectes, com la perspectiva i el nombre auri, del qual es deriva la regla dels dos terços aplicada al cinema i a la fotografia.

c) Espais i escenaris. El seu significat gira entorn de dos paràmetres: l'ésser humà condicionat pel seu ambient natural, i la seva capacitat per a transformar aquest ambient i crear escenaris artificials. D'altra banda, el documentalista també haurà de considerar l'espai de les relacions interpersonals i tota la seva simbologia (posició dels braços i cames, posició frontal o inclinada, distància, etc.).

d) Objectes. Els objectes com a tal són portadors de missatges que tenen a veure tant amb la identitat cultural de les persones (segons el seu disseny, podem detectar la comunitat a la qual pertanyen i quin és el seu lloc dins d'aquesta), com amb els seus valors individuals (per mitjà de la decoració dels espais les persones emeten determinats missatges). Una altra de les propietats que tenen els objectes és que allarguen la durada del missatge en el temps, a diferència dels missatges verbals i gestuals.

e) Imatge humana. El cos també és un potent instrument de comunicació. L'anatomia, la fisonomia i la indumentària ens transmetran importants missatges comunicatius. Així doncs, els trets del rostre i del cos, i també les postures, els vestits i els complements, podran transmetre també trets del caràcter o de l'estat d'ànim de les persones.

D'altra banda, també s'han de tenir en compte un conjunt d'**elements narratius** (accions) (Pinto Molina i altres, 2002):

a) Aspectes sistèmics de la narració visual. La narració visual intenta transmetre percepcions, sensacions, emocions i símbols dins d'uns límits temporals molt concrets. Per això es recorre a processos que succeeixen en la comunicació escrita i oral: descomposició de la història per mitjà d'el·lipsis, plans, preses, escenes i seqüències, i recomposició mitjançant el·lipsis, transicions i continuïtats. L'ambientació i la composició ajudaran a crear sensacions de moviment, ritme i tensió.

b) Activitat expressiva (gestualitat). Els moviments estereotipats i els gestos i les postures resultants transmeten també missatges comunicatius molt potents. L'anàlisi dels gestos es pot fer del més general al més particular. Així, es presta atenció primer als gestos del cos en el seu conjunt, i es continua pels braços, les mans, el rostre i, finalment, la mirada.

c) **Activitat instrumental i objectes.** L'anàlisi se centra en les activitats econòmiques, socials, polítiques, lúdiques, etc. En l'àmbit multimèdia, aquestes activitats es fan servir generalment com a ambientació o pretext. D'altra banda, les activitats instrumentals que provoquen resultats inusuals, i també el comportament dels objectes, requereix una atenció especial dels documentalistes.

1.3.2. L'anàlisi del so

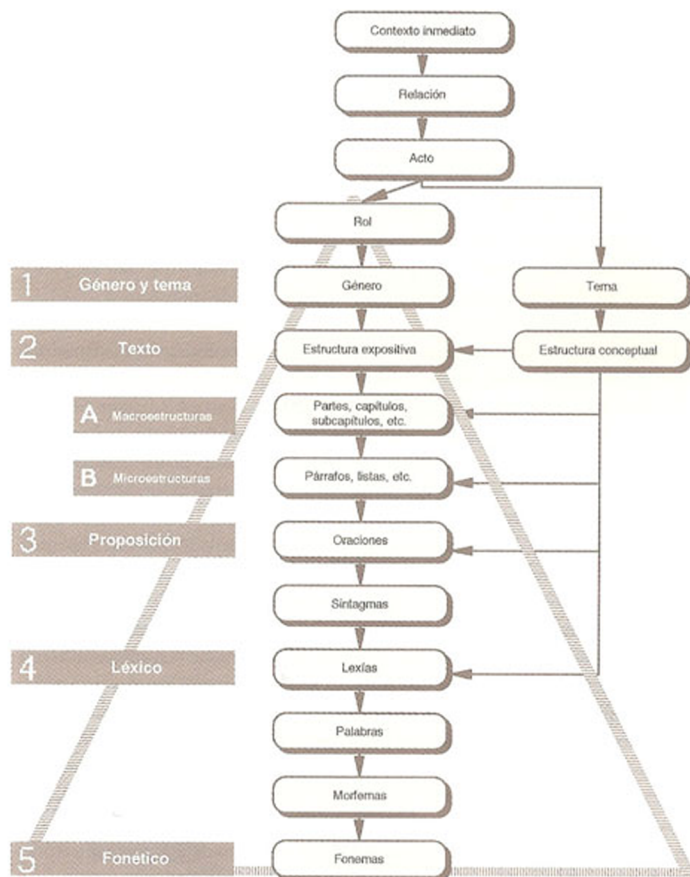
Durant el procés d'anàlisi del so cal tenir en compte, d'una banda, el **soroll**, el **so** i el **silenci** i, de l'altra, la **veu humana**. És important que durant l'anàlisi de la banda sonora es distingeixin clarament els sons i els sorolls. Els sons tenen diverses qualitats que transmetran missatges comunicatius i poderosos efectes emocionals: intensitat, durada, to, timbre i ritme. D'altra banda, la música és la forma no verbal més depurada de comunicació mitjançant sons. En els documents multimèdia s'utilitza per si mateixa o com a acompanyament i té la capacitat de potenciar efectes dramàtics, i crear un clima emocional, social i cultural. Finalment, no hem d'oblidar que el silenci també és un potent recurs comunicatiu.

La veu humana com a llenguatge verbal també és objecte d'anàlisi dels documentalistes, juntament amb els seus codis d'acompanyament, com són les característiques, la intensitat, el to, el timbre, el ritme, l'arrossegament, la ressonància, la col·locació, l'entonació, els sons anatòmics i emocionals (riure, plor, sospir, badall...), els símbols orals (onomatopeies, xiulets...), els reguladors de la conversa, etc.

1.3.3. L'anàlisi verbal

Com ja sabem, el llenguatge verbal és un sistema complex que funciona amb un sistema de nivells organitzats jeràrquicament. Els nivells d'anàlisi es resumeixen en la figura següent:

Figura 2. Nivells d'anàlisi del llenguatge verbal



Font: Pinto Molina i altres (2002)

1.3.4. La interactivitat

La interactivitat està representada en els entorns virtuals, molt importants tant en l'ensenyament i l'oci com en l'àmbit artístic. Dins del procés d'anàlisi documental del contingut, els resums han d'indicar les possibilitats d'interacció que permeten els programes, i els descriptors el tipus de document interactiu que es tracti, representant de manera precisa aquestes possibilitats.

Lectura complementària

Si voleu saber com es fa una anàlisi documental detallada de fotografies, documents audiovisuals televisius, documents videogràfics i cinematogràfics, i recursos electrònics (CD-ROM, webs institucionals, entorns virtuals d'aprenentatge, bases de dades, llibres virtuals i programes d'ordinador), llegiu els estudis de cas que proposa el llibre següent:

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

1.4. Eines per a la descripció documental: tesaurs, taxonomies, ontologies i topic maps

Els professionals de la documentació utilitzen diverses eines per a dur a terme la complexa tasca de descripció dels continguts multimèdia. Aquestes eines són les següents:

a) Els tesaurus

Sovint es comparen amb els diccionaris, però, tot i que contenen unitats lèxiques, tenen diferències molt clares. Els diccionaris contenen diverses definicions d'una unitat lèxica en un context lingüístic, mentre que els tesaurus contenen termes d'indexació i recuperació d'una àrea del coneixement. Ens mostren, per tant, el camp d'aplicació d'un terme i la seva posició en la disciplina. D'altra banda, els podem definir com tres tipus de llenguatge (Naumis Peña, 2007):

- Llenguatge documental, perquè recullen paraules simples i compostes que integren les expressions utilitzades pels autors dels documents d'una àrea temàtica concreta. D'aquesta manera actuen com a termes de recuperació per a aquell qui busqui informació sobre continguts temàtics.
- Llenguatge d'indexació, perquè permeten representar els temes dels documents i els missatges implícits d'aquests. Per tant, asseguren el tractament del contingut temàtic del document.
- Llenguatge controlat, perquè inclouen termes validats per un procés en el qual s'estableixen relacions entre aquests per tal de mantenir la univocitat dels seus descriptors i es determina la preferència d'un terme per damunt dels altres possibles.

La classificació de la informació en un tesaurus es fa per mitjà de sistemes alfabètics, numèrics, alfanumèrics, facetats o temàtics que serveixen per a agrupar els documents en un ordre classificatori després de sotmetre'ls al procés d'indexació. De la mateixa manera que els vocabularis controlats o les llistes d'encapçalaments, els tesaurus normalitzen els temes que contenen els documents. A diferència de la indexació per paraules clau, la informació indexada a partir de termes normalitzats és molt més precisa amb vista a la recuperació, ja que informa dels termes que no estan acceptats i, per tant, no han de ser utilitzats, i també de totes les temàtiques que pot contenir el document sense que hagin estat expressades de manera implícita en el contingut d'aquest.

Figura 3. Resultat de la cerca feta sota el terme *artista* en el tesaurus de la UNESCO.


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

Thesauro 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** ⊖

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

Enllaç recomanat

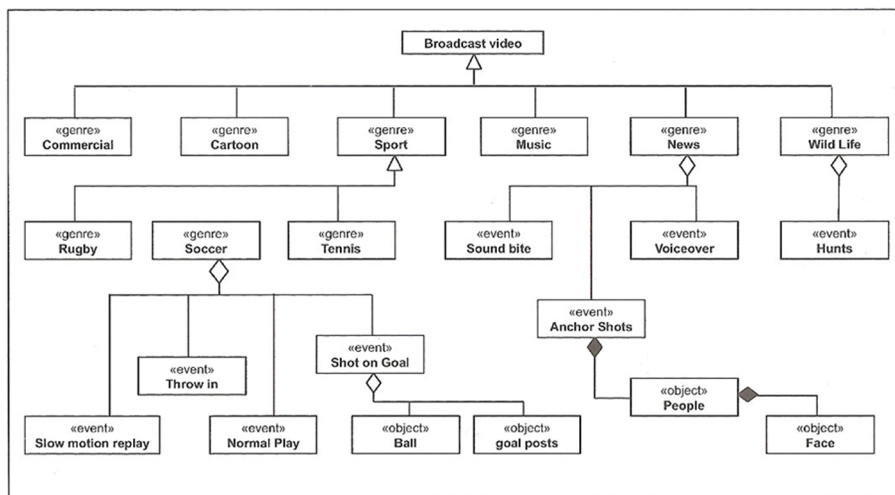
Si voleu conèixer més a fons l'estructura i el funcionament d'un tesaurus, consulteu el tesaurus de la UNESCO, que conté milers de termes controlats relacionats amb l'àmbit de l'educació, la cultura, les ciències naturals, les ciències socials i humanes, la comunicació i la informació: <http://databases.unesco.org/thessp>

b) Les taxonomies

Es tracta d'un conjunt de termes organitzats de manera jeràrquica i a la vegada classificats per categories i subcategories, de manera que cada terme està relacionat amb els termes dels quals depèn i amb el termes que en depenen. Així doncs, les taxonomies presenten relacions semàntiques dels termes de manera més àmplia que els tesaurus, ja que poden incloure dades corporatives en els termes i en la megaestructura mateixa on s'afegeixen, a més, diversos tesaurus. En definitiva, poden abraçar tota la informació sobre una organització/institució, és a dir, els seus processos de negoci, la normativa i pautes sobre la manera de treballar del personal, els continguts que tracta, etc.

Normalment, les empreses del sector audiovisual acostumen a utilitzar per al seu treball intern un conjunt de classificacions jeràrquiques aplicades, per exemple, als tipus de programes que produeixen, als personatges que hi prenen part o als tipus de formats de transmissió tècnica dels continguts. Malauradament, no acostumen a relacionar cada una d'aquestes classificacions entre si per formar una taxonomia única, ja que és una tasca força costosa. El cert és que, de moment, no hi ha una taxonomia ben definida i acceptada per als objectes digitals actuals. És un tema que els experts estan debatent a fons.

Figura 4. Extracte d'una taxonomia



Font: Roach i altres (2002)

Enllaços recomanats

Si voleu conèixer més a fons el funcionament d'una taxonomia, consulteu les *NewsCodes*, creades per l'International Press Telecommunications Council (IPTC). Les agències de premsa i els mitjans de comunicació les fan servir per a l'assignació de metadades sobre els tipus de gèneres periodístics, temes, plans d'imatges fixes o vídeo, etc.: <http://www.iptc.org/site/NewsCodes/Overview>

Llegiu aquest blog interessant on s'expliquen, des d'un punt de vista particular, aspectes relacionats amb l'ús de les taxonomies en les organitzacions, en les xarxes socials, etc.: <http://thetaxonomyblog.com>

Consulteu el web del Taxonomy Boot Camp de Washington, un congrés anual en què es mostren els últims avenços en organització, gestió i integració d'informació i dades en les empreses. Està dirigit especialment als professionals de la informació, com ara gestors de continguts, especialistes en taxonomies i classificacions, arquitectes de la informació, enginyers del coneixement, creadors de portals i intranets, etc.: <http://www.taxonomybootcamp.com>

c) Les ontologies

Les ontologies defineixen els possibles tipus de continguts existents en un fons documental digital, és a dir, les seves característiques i elements, i també les diferents relacions que s'estableixen entre aquests. No totes aquestes relacions han de ser aplicades a tots els elements i característiques, de la mateixa manera que un mateix element o característica pot tenir més d'una relació.

Val a dir que aquestes eines de descripció documental són força complexes de dissenyar, ja que defineixen una gran quantitat i varietat de relacions. És per això que són utilitzades, en la majoria dels casos, per grans institucions. Exemples en serien la BBC i la European Broadcasting Union (EBU).

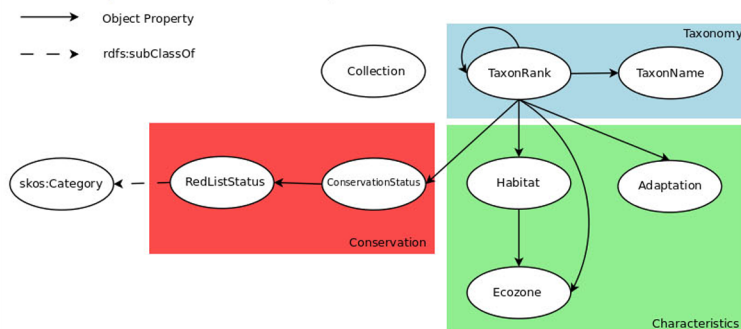
A la Xarxa, el concepte d'ontologia està estretament lligat al Web semàntic, que pretén convertir els recursos que es troben disponibles a la Xarxa en una gran font d'informació, accessible per mitjà de tot tipus d'eines automatitzades. Podríem dir que les ontologies compleixen funcions de diccionaris per a etiquetar el llenguatge web amb precisió, establir-ne les propietats i les relaci-

ons amb altres conceptes i transformar-los en termes, per tal que les eines de cerca interpretin el significat de les paraules trobades. Són, per tant, diferents dels motors de cerca, que funcionen amb paraules clau.

Figura 5. Diagrama que il·lustra les relacions entre les classes clau a la Wildlife Ontology, creada per la BBC

Vocabulary Diagram

The following diagram illustrates the relationships between the key classes in the ontology. A number of classes, e.g. sub classes of TaxonRank, Habitat and Adaptation have been omitted for clarity.



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](#) | [AnimalIntelligence](#) | [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](#)

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

Enllaç recomanat

Per a saber com funciona una ontologia, consulteu el web de la Wildlife Ontology, desenvolupada per la BBC: <http://www.bbc.co.uk/ontologies/wildlife/2010-02-22.shtml>

Lectura complementària

Per a entendre les diferències entre el funcionament d'un tesaurus i una ontologia, llegiu l'article següent:

L. Codina; R. Pedraza-Jiménez (2011, setembre-octubre). "Tesauros y ontologías en sistemas de información documental". *El profesional de la información* (vol. 20, núm. 5, pàg. 555-563). http://www.lluiscodina.com/ontologiaTesauros_2011.pdf

d) Els *topic maps* (mapes temàtics)

Es tracta d'un estàndard que s'està implantant en el Web i que possibilita la navegació conceptual. Estan relacionats amb el Web semàntic i es desenvolupen en llenguatges com XML i XTM. Els termes que s'extreuen durant el procés d'indexació dels documents tenen relació amb un terme genèric i termes associats, igual que els tesaurus. La ubicació en l'àmbit de coneixement estarà definida per les relacions que tenen amb els termes que es troben en el document mateix. Ara bé, aquests termes no seran únics, com passa amb els tesaurus, sinó que es faran servir tots els sinònims que apareguin en el document i variaran en funció del llenguatge utilitzat per cada autor. En definitiva, els *topic maps* no són vocabularis definits prèviament, sinó que són un model flexible amb el vocabulari del document indexat mateix. D'altra banda, val a dir que tant els tesaurus com les taxonomies i les ontologies poden ser una font molt valuosa per a la construcció dels *topic maps* (Naumis Peña, 2007).

Lectura complementària

Si voleu saber més sobre els *topic maps*, llegiu l'article:

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

1.5. El concepte de metadades

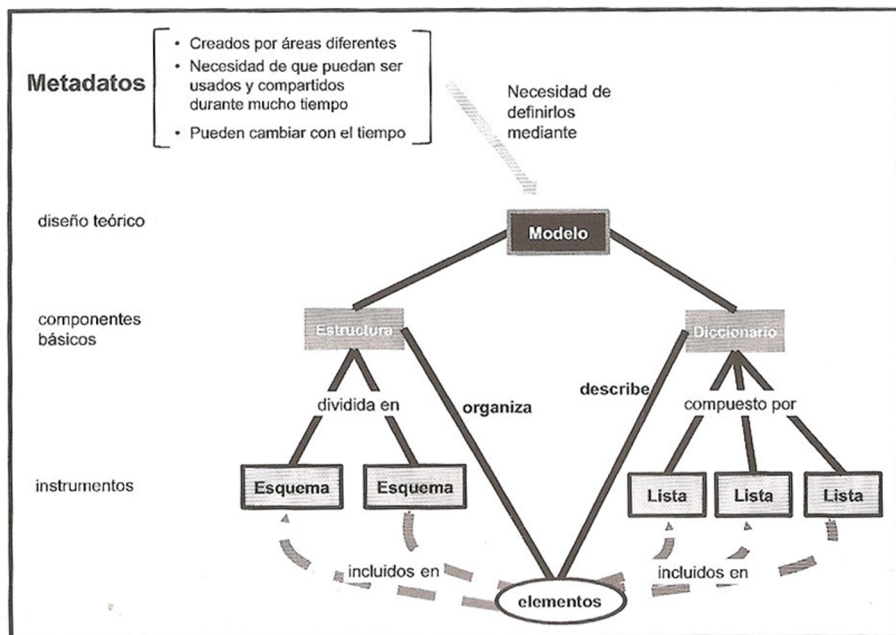
Com ja hem comentat a l'inici d'aquest apartat, les metadades són l'element mínim que s'utilitza per a identificar i descriure continguts digitals. Es tracta d'un conjunt de dades que aporten informació sobre cada document. En el cas d'un document multimèdia, seria el conjunt de dades que identifiquen i descriuen la combinació de vídeo, àudio, text, gràfics, i que giren entorn de les característiques físiques, la durada de tot el conjunt, el lloc de realització, les descripcions del títol o el tema, etc. S'utilitzen en qualsevol moment del flux de treball, és a dir, es poden incorporar cada vegada que es creï, es modifiqui, es relacioni, s'emeti, s'arxivi o es recuperi un contingut.

Les metadades han de complir uns aspectes mínims (Caridad i altres, 2011):

- Cadascuna de les àrees de treball en una producció multimèdia voldrà fer servir aquelles metadades que els siguin més útils. Els professionals (enginyers) encarregats de transmetre els continguts han de tenir alguns requisits per a empaquetar-les i facilitar-ne així la difusió.
- Sovint passa que aquests continguts s'han d'utilitzar i compartir durant molt de temps. Això implica, per tant, que les metadades han de ser dissenyades perquè es mantinguin en un futur, no solament durant l'ús concret que se'n pugui fer en un determinat moment. A més, han de ser útils per a múltiples usuaris.
- Tant els continguts com les metadades ja afegides han de permetre que altres usuaris/documentalistes puguin anar afegint noves metadades per a descriure el tema i els personatges amb més profunditat, etc.
- En un futur, aquestes metadades s'hauran de millorar o canviar.

Així doncs, el valor de les metadades es troba en la seva relació amb altres metadades, és a dir, com s'han de definir de manera que no es dupliquin en les diferents àrees de treball, en quin ordre han de ser introduïdes, quin ha de ser el seu grau de complexitat, si han de contenir llenguatge controlat o lliure, etc. En definitiva, les metadades que s'utilitzin haurien de tenir els components següents:

Figura 6. Components bàsics d'un model de metadades



Font: M. Caridad i altres, 2011

El cert és que tot el conjunt de metadades d'un sistema multimèdia s'hauria de recollir en un model teòric que especifiqui tots els requisits necessaris per a utilitzar-lo. D'una banda, s'haurien d'organitzar en una estructura comuna, ja que sovint són utilitzades per diversos usuaris. De l'altra, tots els elements que conformen les metadades haurien d'estar recollits en un diccionari que permeti identificar el nom assignat a cada element de manera clara, definir per què s'utilitza aquest element i unificar criteris per evitar que un nom concret s'utilitzi per a elements diferents.

1.6. Estructures de metadades

Qualsevol estructura de metadades hauria de complir els requisits següents (Rust i Bide, 2000):

a) **Identificació única:** cada contingut ha d'incloure almenys un identificador, que a la vegada ha de complir una sèrie de condicions:

- Ha de ser estable, és a dir, no pot ser modificat. Ara bé, pot ser un identificador relatiu. Això significa que si un objecte canvia físicament d'ubicació (per exemple, dins d'una base de dades accessible per mitjà d'Internet), els usuaris poden utilitzar el mateix identificador per a accedir al contingut, mitjançant un sistema d'adreçaments.
- Ha de ser únic dins un sistema d'identificació. Per exemple, si un contingut està format per àudio i vídeo, es podrà utilitzar un identificador per a tot el conjunt, un altre per a l'àudio, i un altre per al vídeo.
- Ha de ser segur, és a dir, s'han d'utilitzar algorismes i tècniques de xifratge, perquè els programes que facin servir els identificadors facin comprovacions digitals i els donin per vàlids.

- Ha de proporcionar accés a metadades bàsiques que serveixin per a descriure el contingut identificat.

b) Assignació d'autoritat: això significa que cada vegada que s'introdueixi o es modifiqui una metadada s'haurà d'identificar clarament l'autor.

c) Granularitat funcional: en les produccions multimèdia és molt útil identificar tant el contingut en el seu conjunt com cadascun dels possibles fragments. Això es fa des de diferents punts de vista: tècnic (seqüències, plans), semàntic (talls separats per tema o personatge), temporal (detectar el començament de cada inici de fragment dins d'un programa determinat, per exemple), entre d'altres.

d) Permís d'accés: és important donar accés a les metadades que necessitin els usuaris, però també s'ha de restringir tot el personal que no estigui autoritzat, bé perquè no tinguin els coneixements tècnics suficients, bé perquè siguin dades d'entrada restringides.

1.7. Creació i emmagatzematge de metadades

Durant el procés de creació multimèdia és preferible que la inserció de les metadades interfereixi el menys possible. Actualment moltes eines de filmació i gravació ja inclouen la possibilitat d'afegir automàticament metadades relacionades amb la data de la presa, la durada, el codi de temps de cada imatge captada, la localització geogràfica, etc. D'altra banda, molts sistemes de gestió de continguts també les generen automàticament (per exemple, reconeixent elements que apareixen repetidament en diferents imatges, el grau de lluminositat que permet distingir la localització i el moment del dia, etc.). Finalment, hi ha maneres d'afegir metadades de manera semiautomàtica, escollint els termes entre una sèrie tancada d'opcions definida prèviament per mitjà d'un vocabulari controlat.

L'emmagatzematge d'aquestes metadades també és un procés important que cal tenir en compte. Es fa de dues maneres:

- Desant en un mateix suport físic tot el contingut, és a dir, tant el material com les metadades. Serien les anomenades metadades incrustades o *embedded*. És l'opció més adequada i és molt útil, sobretot quan les metadades no canviaran en el temps.
- Desant per separat el material i les metadades, mantenint-ne la relació. En aquest cas parlem de metadades externes. Aquest sistema és útil per a metadades descriptives o que s'hagin de modificar sovint.

Sigui com sigui, l'important és mantenir el codi que uneix cada essència (és a dir, qualsevol dada necessària per a representar una experiència sensorial, visual, oral o de qualsevol altre tipus), amb les seves metadades respectives. Si això no funciona, es farà difícil localitzar i reconèixer el contingut que s'està buscant. Aquest codi s'estableix de dues maneres (Caridad i altres, 2011):

- Mitjançant XML, el llenguatge de marcatge de text que constitueix una sèrie de regles bàsiques que faciliten la definició d'estructures de metadades i les metadades basades en aquestes estructures.
- Mitjançant el format KLV (*Key, Length, Value*), enfocat principalment per a la transmissió de contingut. Cada metadada es troba dins un camp numèric, del qual es descriuen tres característiques: la clau (*key*), que identifica el contingut; la longitud (*length*) numèrica del camp, i el valor (*value*) que es dona per a aquest camp.

1.8. Tipologia de metadades

Hi ha diversos tipus de metadades:

a) Metadades físiques, que proporcionen informació sobre la ubicació dels fitxers, les formes d'emmagatzematge, dades sobre autenticitat i seguretat, els formats de digitalització, les ràtios de compressió, la velocitat de reproducció, els tipus de descodificadors de reproducció, la resolució, la freqüència de mostreig, el format de sortida, la quantitat de pistes d'àudio, etc.

b) Metadades administratives, utilitzades per a la gestió i administració dels continguts digitals, proporcionen informació sobre el mètode d'adquisició, els drets, els criteris de selecció del contingut, el control de les diferents versions, els usuaris interns que hi han accedit, etc. També ofereixen informació sobre les condicions d'ús, si es tracta d'un original o una còpia, quines característiques materials i informàtiques té (suports per a la reproducció, per exemple), etc.

c) Metadades descriptives, utilitzades tradicionalment pels documentalistes, informen sobre l'autor, el títol, la durada, la data de producció, el lloc d'edició, la ubicació física on es va fer l'enregistrament o la gravació, etc. També s'hi afegixen els descriptors temàtics, tant paraules clau en llenguatge lliure com descriptors extrets de llenguatges documentals controlats (tesaurus, llistes jerarquitzades de termes, taxonomies, etc.), i també transcripcions d'àudio, entre altres dades.

2. La preservació de continguts digitals multimèdia

2.1. Què entenem per preservació digital?

Com sabem, la informació en paper i altres suports, com per exemple els microfilms, pot durar centenars i milers d'anys, en alguns casos. En canvi, la informació codificada en format digital és bastant improbable que duri més d'una dècada o dues. Les raons principals són diverses: l'envelliment dels suports digitals, l'obsolescència dels formats, del programari i del maquinari, i els problemes d'incompatibilitat dels nous sistemes (Bia i Sánchez, 2002).

Definim la *preservació digital* com la gestió necessària per a assegurar el manteniment i l'accés continu als materials creats en format digital (coneguts també com a *born digital*) a llarg termini.

Per garantir aquest accés, caldrà tenir en compte el següent (Keefer i Gallart, 2005):

- L'evolució de la tecnologia.
- La coordinació de totes les institucions implicades (*stakeholders*) durant el procés de creació, distribució, emmagatzematge i ús d'aquests continguts. Aquestes institucions van des de les empreses i les administracions públiques fins a les associacions professionals, les biblioteques, els arxius, etc.
- El desenvolupament i l'aplicació de normatives adequades per a dur a terme el procés.

Quan parlem de preservació digital no ens referim al procés de digitalització de continguts ni a fer còpies de seguretat. Aquestes últimes es duen a terme, bàsicament, per protegir els documents davant pèrdues de dades per culpa d'apagades o per trencaments de suports. El que es desa és tant la informació publicada en el servidor de la institució o empresa com els recursos digitals en procés d'edició. La preservació digital, en canvi, s'ocupa de salvaguardar els recursos digitals que necessitem en un futur.

Un exemple que permet veure aquesta diferència és el cas de les imatges gràfiques. D'una banda, interessa fer còpies de seguretat de les que estan en format JPG de baixa qualitat desades en el servidor, i de l'altra, interessa fer còpies de preservació de les que estan en format TIFF d'alta qualitat que no es troben en el servidor (pel seu pes i per la lentitud de transmissió), però que segurament

⁽¹⁾Les tècniques emprades, en aquest segon cas, les veurem en l'apartat 2.4.

es necessitaran per a usos futurs. En definitiva, hi ha diferències notables entre els dos processos. Si bé les còpies de seguretat es fan diàriament o setmanalment, les còpies de preservació es fan normalment cada any o any i mig¹ (Bia i Sánchez, 2002).

2.2. Els repositoris digitals fiables (*trustworthy digital repositories*)

Els repositoris digitals fiables són institucions que tenen com a objectiu preservar a llarg termini els recursos digitals que custodien. Aquestes institucions van des de biblioteques i arxius nacionals fins a universitats i altres organismes acadèmics.

Els repositoris digitals fiables són responsables del següent:

- Mantenir la infraestructura tècnica adequada.
- Garantir la integritat i l'autenticitat dels recursos.
- Vigilar la protecció dels drets intel·lectuals.
- Limitar-ne l'accés només als usuaris autoritzats.
- Mantenir les eines d'identificació i recuperació de la informació.
- Subministrar les metadades normalitzades i actualitzades que han d'acompanyar la informació.

Hi ha iniciatives que avaluen aquests repositoris. Es tracta d'auditories i certificacions que ofereixen sistemes per a mesurar l'adequació i precisió dels processos que duen a terme (plans de formació, currículums del personal, descripcions de treball, etc.). L'objectiu és analitzar la infraestructura organitzativa, la gestió dels continguts digitals, la tecnologia, la infraestructura tècnica i la seguretat que ofereixen aquestes institucions. Algunes d'aquestes certificacions són **TRAC** (*Trustworthy Repositories Audit & Certification*, <http://www.crl.edu/pdf/rac.pdf>), **NESTOR** (*Catalogue of Criteria for Trusted Digital Repositories*, <http://edoc.hu-berlin.de/series/nesstor-materialien/8en/PDF/8en.pdf>) i **DRAMBORA** (*Digital Repository Audit Method Based on Risk Assessment*, <http://www.repositoryaudit.eu>).

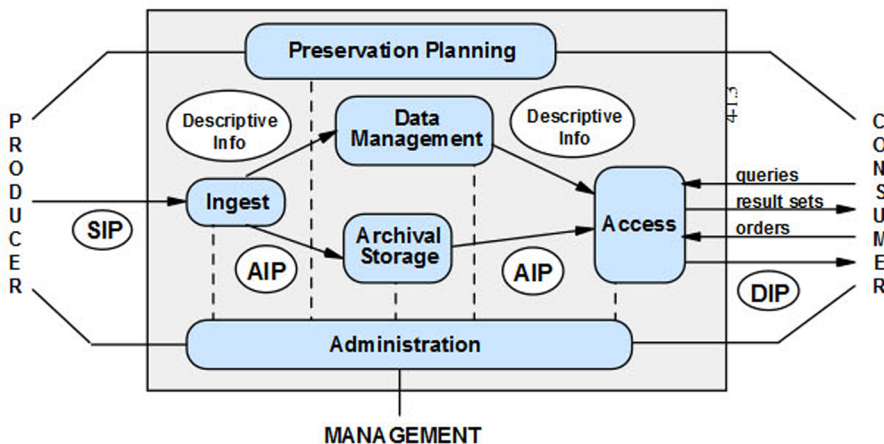
2.3. Funcions d'un repositori digital fiable: el model OAIS

L'any 2002, la NASA va publicar l'**Open Archival Information System (OAIS)** (<http://ddp.nist.gov/refs/oais.pdf>), un model acceptat i estandarditzat com a norma ISO 14721, que estableix les funcions que ha de complir un sistema de preservació digital, i també la manera de gestionar-ne els fitxers i metadades. Aquestes funcions són bàsicament sis:

- Ingesta: entrada de dades i fitxers. És un dels mòduls més desenvolupats dels sistemes de preservació. Per mitjà de polítiques coherents, els productors multimèdia haurien de crear fitxers en formats estàndard.
- Emmagatzematge: gravació de dades.

- Gestió de dades: introducció de les metadades corresponents.
- Accés: mòduls de consulta.
- Preservació: duta a terme per la direcció interna de la institució encarregada de la preservació. Tindrà molta importància la vigilància tecnològica.
- Serveis comuns: comprovació interna, perquè tot funcioni correctament.

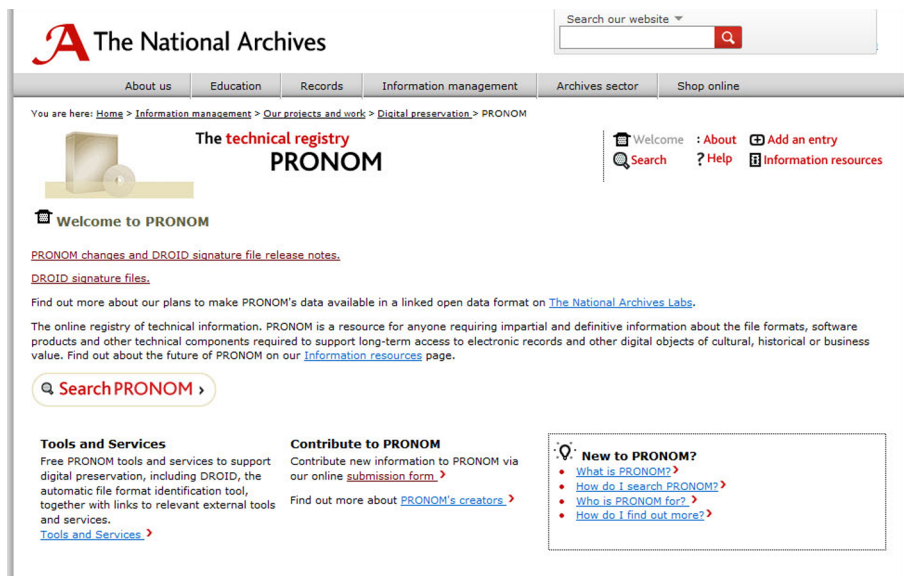
Figura 7. Diagrama on es mostren les sis funcions del model OAIS, creat per la NASA.



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

Una de les funcions que pot presentar més problemes és la ingesta de dades. De vegades, els fitxers que s’han d’introduir en un sistema de preservació digital tenen extensions il·legibles o que es fan difícils de reconèixer. La solució implica utilitzar programes de detecció de formats que contenen, a més, el registre de tots els formats existents a escala mundial. Un exemple és PRONOM (<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>), desenvolupat per The National Archives del Regne Unit.

Figura 8. Portada de PRONOM



D’altres programes que identifiquen i validen formats són els següents:

- **DROID** (*Digital Record Object Identification*, <http://sourceforge.net/projects/droid>): gratuït i també creat per The National Archives del Regne Unit.
- **JHOVE** (*JSTOR/Harvard Object Validation Environment*, <http://jhove.sourceforge.net>): orientat a sistemes més professionals.
- **XENA** (<http://xena.sourceforge.net>): gratuït, desenvolupat per The National Archives of Australia i basat en Java.

2.4. Estratègies tècniques de preservació digital

Els principals implicats en el tema de la preservació dels fons documentals són, tradicionalment, les biblioteques i els arxius. Ara bé, els productors, en aquest cas multimèdia, també poden facilitar aquesta tasca des del moment en què dissenyen i creen els continguts, posant un èmfasi especial en els formats, la consistència, la identificació i descripció, i la normalització. Per tant, és necessari tenir en compte qüestions de conservació en totes les etapes del cicle de vida dels documents.

En aquest sentit, les empreses i institucions haurien de dissenyar un pla de preservació que respongui les preguntes següents (Bia i Sánchez, 2002):

- Què s'ha de conservar i per què?
- On ho hem de conservar?
- Fins quan cal conservar-ho?
- Com ho trobem després?
- Com aconseguim que es mantingui inalterat?
- Com evitem que es torni obsolet?

És important que aquest pla sigui revisat de manera periòdica i assignar als continguts preservats un límit de vida. D'altra banda, els implicats en aquest procés han de prendre consciència que cap dels suports digitals actuals no garanteix una vida llarga de la informació que contenen. Aquesta informació està constantment exposada a una sèrie de factors que en fan perillar la perdurabilitat: negligències humanes, errors tècnics i mecànics, virus, obsolescència, incompatibilitat i pèrdua dels programes, canvis no documentats ni autoritzats, metadades incompletes, manca d'actualització de la informació, etc.

En el moment de dur a terme una política de preservació digital les empreses i institucions n'han d'analitzar a fons la viabilitat. Això significa que hauran de veure si són capaces d'assumir i complir les responsabilitats associades a aquesta política. Malauradament, podem afirmar que encara hi ha una manca generalitzada de finançament, d'assignació de responsabilitats, de polítiques ben dissenyades i de plans de formació per al personal. D'altra banda, també hi ha tota la qüestió legal, encara poc resolta, i que gira entorn del següent:

- El dret de modificació dels continguts, és a dir, el pas d'un format a un altre.

Enllaços recomanats

Alguns dels projectes de referència que es basen en l'elaboració i aplicació del model OAIS són **PLANETS** (<http://www.planets-project.eu>), i **CASPAR** (*Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval*, <http://www.casparpreserves.eu>), tots dos cofundats per la Unió Europea i altres institucions dedicades a la preservació digital.

- El dret de recollir i arxivar els continguts.
- El dret de proporcionar l'accés al públic.

Un cop s'hagi analitzat la viabilitat, l'empresa o institució ha de procedir amb les diferents estratègies de preservació que més convinguin. Algunes d'aquestes són:

a) *Refreshing* o actualització de les dades en un suport nou

Representa una solució al problema del deteriorament dels suports digitals. Es produeix a partir de la transferència de les dades d'un suport a un altre, mantenint intacte el seu *bitstream* original i sense afectar el programari necessari per a llegir-lo.

b) Migració o canvi de format

És una mesura contra l'obsolescència i consisteix a convertir la informació a nous formats. A diferència del *refreshing*, aquesta tècnica sí que altera el *bitstream* original i implica també un canvi de programari per a la lectura. Un dels riscos més importants és la possible pèrdua de dades i de la funcionalitat de l'entorn original. A més, s'ha de tenir en compte que els formats queden obsolets amb el pas del temps.

c) Emulació

Consisteix a recrear l'entorn de programari i maquinari. Permet, per tant, que el programari original sigui utilitzat sense necessitat que el sistema original que l'executava continuï existint. És important preservar tant l'emulador com el sistema operatiu, l'aplicació i les dades. Algunes arquitectures emulades són: Atari2600, Nintendo 64, Game Boy, PlayStation i Sega Genesis.

Com a exemples d'emuladors de videoconsoles tenim el **MAME** (<http://www.mamedev.org>) (emulador de màquines recreatives Arcade), el **CybersteIla** (<http://stella.sourceforge.net>) (emulador de la videoconsola Atari 2600) i el **zSNES** (<http://sourceforge.net/projects/zsnes>) (emulador de la videoconsola Super Nintendo). Un altre bon exemple d'emulador és el **Dioscuri** (<http://dioscuri.sourceforge.net>), desenvolupat per la Koninklijke Bibliotheek i el Nationaal Archief dels Països Baixos.

d) Estandardització

Es tracta de buscar una manera de codificar i formatar els continguts digitals adherint-se a estàndards reconeguts i més utilitzats. Es pressuposa que aquests estàndards perduraran en el temps i aniran solucionant els problemes de compatibilitat que vagin apareixent amb l'evolució de l'entorn informàtic.

Enllaç recomanat

El UK Data Archive treballa amb múltiples formats de dades per a diferents usos. Consulteu aquest enllaç que conté els formats de preservació acceptats per aquest organisme: <http://data-archive.ac.uk/create-manage/format/formats-table>

e) Encapsulació

Consisteix a empaquetar la informació que s'ha de preservar juntament amb les seves metadades descriptives. Per mitjà d'estructures físiques o lògiques, anomenades *containers* o *wrappers*, es proporciona una relació entre tots els components de la informació que cal preservar. Aquesta estratègia pretén evitar l'obsolescència de la informació digital, disminuint la possibilitat que es perdin els elements necessaris per a codificar i traduir les dades, ja que els detalls sobre com cal interpretar els bits en qualsevol sistema futur queden registrats.

Per a l'encapsulació, s'ha creat un format específic anomenat *UPF (Universal Preservation Format)*. El seu objectiu és que els objectes digitals siguin independents del programari i dels sistemes operatius, però que les metadades que portin incorporades incloguin les especificacions tècniques per a tenir accés als materials encapsulats.

f) Ordinador virtual universal (UVC)

Es tracta de la representació virtual d'un ordinador simplificat que permet desar informació actual codificada per a desxifrar-la en un futur. Per a poder funcionar, utilitza elements tant de la migració com de l'encapsulació, tot i que per a molts es considera una tècnica d'emulació. Consisteix a fer que un programa independent registri, des d'un principi, l'arquitectura bàsica de cada ordinador per tal de poder simular-ne les memòries, les seqüències de registres i la manera de connectar-se entre aquests, per a reconstruir els fitxers en el seu aspecte original per mitjà d'un descodificador de formats.

g) Arqueologia digital

Consisteix a recuperar informació a partir de fonts de dades fetes malbé, fragmentades o arcaïques. El procés consisteix a desxifrar el contingut, crear estàndards de migració i copiar en format llegible per als futurs usuaris. La British Library és una de les pioneres en utilitzar aquesta tècnica de preservació.

2.5. Exemples de programari per a la preservació digital

La majoria dels grans sistemes que es troben avui dia en el mercat i que són utilitzats per algunes institucions dedicades a la preservació digital, com ara les biblioteques o els arxius nacionals, preveuen funcions importants com la ingesta de dades, la importació i transformació de metadades, la migració de formats i la implantació de polítiques de vigilància tecnològica. Val a dir, però, que la majoria d'aquests sistemes tenen un cost econòmic força elevat. Ara bé, també s'estan desenvolupant programaris de codi obert, utilitzats tant per empreses privades com per institucions acadèmiques. Vegem-ne alguns exemples:

a) Programaris comercials

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

Desenvolupat per IBM i la Koninklijke Bibliotheek dels Països Baixos.

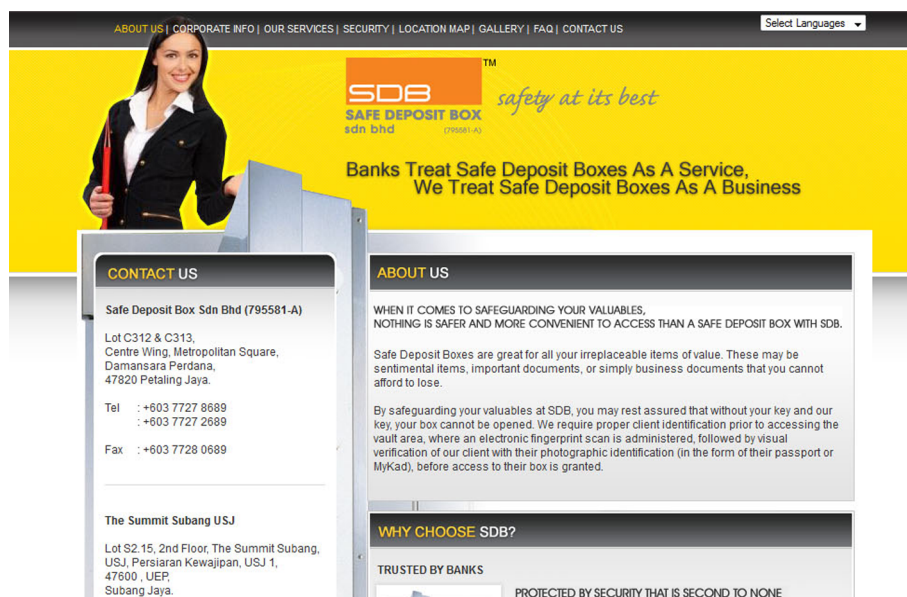
Figura 9. Portada de DAIS



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

Desenvolupat per Tessella, The National Archives del Regne Unit i el National Archives and Record Administration (NARA) dels Estats Units.

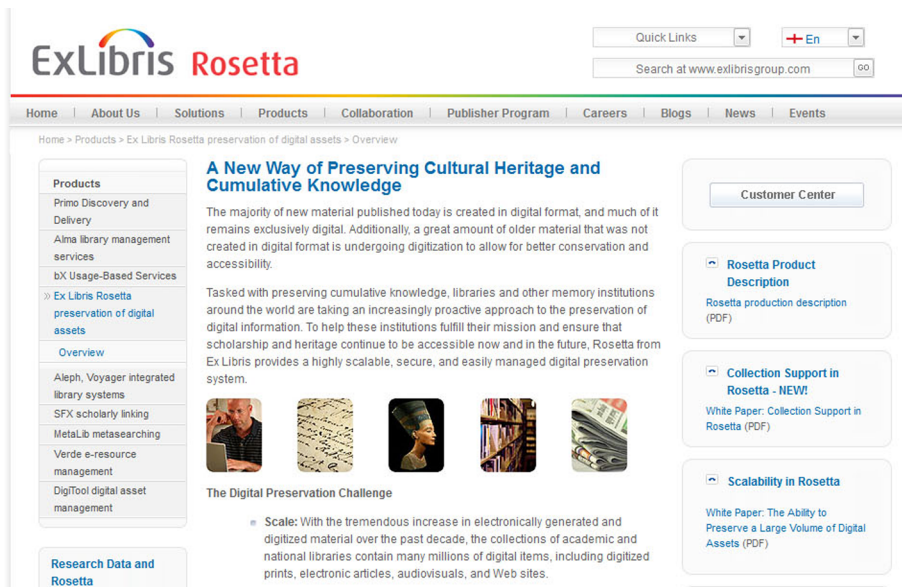
Figura 10. Portada d'SDB



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

Desenvolupat per Exlibris i The National Library of New Zealand.

Figura 11. Portada de Rosetta

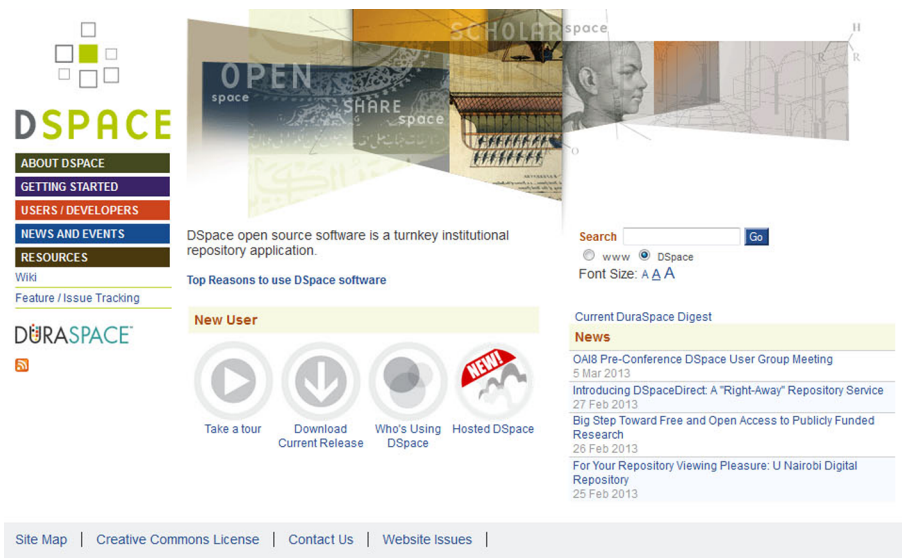


b) Programaris de codi obert

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

Completament gratuït, és utilitzat tant per empreses comercials com per institucions acadèmiques. DuraSpace és l'organització encarregada de desenvolupar-lo i millorar-lo.

Figura 12. Portada de DSpace



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

També d'accés obert, aquest repositori és utilitzat, sobretot, per a preservar arxius digitals personals. DuraSpace és també l'organització encarregada de desenvolupar-lo i millorar-lo.

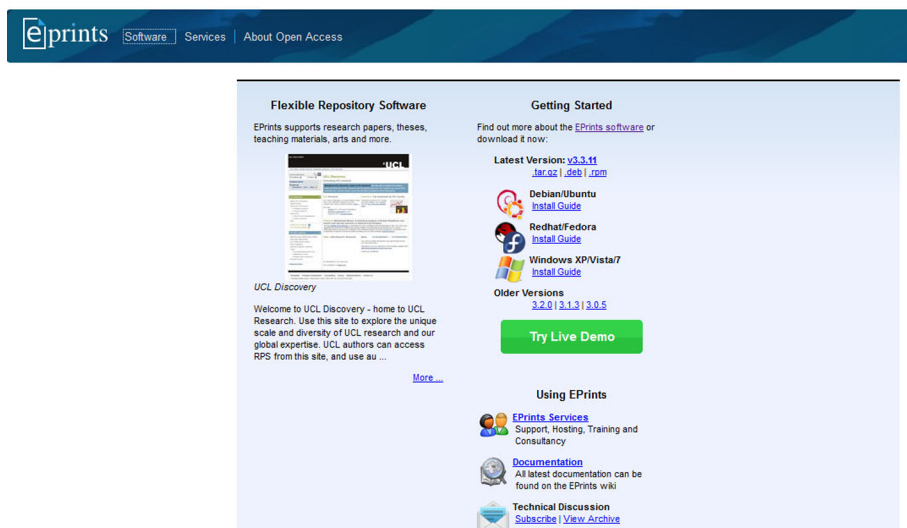
Figura 13. Portada de Fedora



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

També gratuït, garanteix la preservació digital de tot tipus de documents, des de dades científiques, tesis, informes i material per a la investigació, fins a multimèdia.

Figura 14. Portada d'ePrints



2.6. La preservació digital en altres àmbits d'actuació

A més dels grans projectes de preservació digital orientats a grans biblioteques i arxius, s'estan duent a terme investigacions que s'ocupen dels repositoris *open access* i de la preservació lligada als arxius administratius, tant de les empreses privades com públiques. En aquestes empreses, no solament és necessari preservar la documentació que generen, sinó també les firmes electròniques i les marques de temps (*data stamp*). Alguns sistemes que garanteixen aquesta preservació són el català **iArxiu** (<http://www.aoc.cat/Inici/SERVEIS/Gestio-interna/iArxiu>) i l'alemany **ArchSafe** (<http://www.archisafe.de/s/archisafe/index>) (Térmens, 2009).

Pel que fa al sector artístic, un dels comitès de l'International Council of Museums (ICOM), l'International Committee for Museum Documentation (Cidoc), està coordinant la investigació per trobar solucions de preservació dels museus, sobretot pel que fa al *Net.Art* i la perdurabilitat dels sistemes virtuals, com poden ser les reconstruccions virtuals de monuments històrics. Una altra iniciativa que investiga la preservació relacionada amb els museus i l'art multimèdia és el projecte **Documentation and Conservation of Media Arts Heritage (DOCAM)**, coordinat pel Centre for Research and Documentation (CR+D) de la Fondation Daniel Langlois.

Altres àmbits en què s'està investigant són l'aeronàutica, per a consolidar formats, processos i sistemes que assegurin la perdurabilitat de tota la documentació tècnica dels avions; la ciència, en què s'estan elaborant importants bases de dades sobre astronomia, meteorologia, biologia, genètica, etc.; i els arxius personals, com ara els dels partits polítics.

2.7. Polítiques de preservació digital pioneres arreu del món

Fins ara, algunes biblioteques i arxius nacionals, i també universitats, empreses de programari i institucions dipositàries de documentació han liderat la investigació en preservació digital. Els Estats Units, Austràlia, Nova Zelanda i països europeus com els Països Baixos, el Regne Unit, França i Alemanya han experimentat els avenços més importants. A Espanya, en canvi, no hi ha hagut massa iniciatives pel que fa aquest tema. No és fins fa ben poc que s'han començat a dur a terme alguns projectes competitiu a escala internacional.

Vegem algunes institucions i iniciatives pioneres (Térmens, 2009):

a) Estats Units

- 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 i University of California)
- DAITSS (Florida Center for Library Automation)
- Jhove (Harvard University)
- Unified Digital Formats Registry (Harvard University)
- Internet Archive
- Portico

b) Austràlia i Nova Zelanda

Enllaç recomanat

Si voleu saber més sobre el projecte DOCAM, escolteu l'àudio de la conferència que va dur a terme Alain Depocas, director del CR+D de la Fondation Daniel Langlois, el 31 de maig del 2010 a l'Auditori del Museu d'Art Contemporani 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>

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

c) Països Baixos

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

d) Regne Unit

- Digital Lives (British Library)
- Join Information Systems Committee (JISC)
- The National Archives
- Digital Curation Centre (CCC) d'Edimburg

e) França

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

f) Alemanya

- Kopal (Deutsche Nationalbibliothek i Niedersächsische Staats-und Universitätsbibliothek Goettingen)
- Archisafe (Physikalisch-Technische Bundesanstalt)

g) Espanya

- Biblioteca Nacional de España (forma part del grup de treball europeu Long Term Archiving and Retrieval)
- Grupo de Trabajo de Patrimonio Digital del Ministerio de Cultura
- Red de Bibliotecas Universitarias Españolas (grup de treball sobre preservació digital en les biblioteques universitàries)
- 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 complementària

Si voleu conèixer més sobre el projecte Patrimoni Digital de Catalunya (PADICAT), dedicat a la preservació del Web i equiparable a les millors iniciatives internacionals, llegiu l'article següent:

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

En definitiva, la investigació en preservació digital encara és un camp emergent i poc definit. En la pràctica, encara hi ha un grau d'implementació força baix, en el sentit que només algunes grans institucions i empreses privades han començat a aplicar sistemes de preservació digital, uns sistemes que encara no s'estan comercialitzant de manera àmplia. La bona notícia és que cada vegada hi ha més coneixement sobre aquest tema, i això fa que vagin proliferant força sistemes de codi obert basats en XML i Java, compatibles amb la majoria d'entorns de treball. D'altra banda, no podem obviar que al nostre país hi ha un cert retard en aquest àmbit, i les iniciatives de biblioteques, arxius, universitats i museus són encara insuficients. S'espera però, que en un futur no gaire llunyà la situació canviï, sobretot pel que fa a les biblioteques nacionals i universitàries, amb recursos digitals cada vegada més voluminosos (Térmens, 2009).

Bibliografia

Bibliografia bàsica

Bía, A.; Sánchez, M. (2002). *Desarrollo de una política de preservación digital: tecnología, planificación y perseverancia*. Alacant: 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*. Mèxic / Buenos Aires: Universidad Autónoma de México / 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, novembre-desembre). "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.

Bibliografia complementària

Codina, L.; Pedraza-Jiménez, R. (2011, setembre-octubre). "Tesauros y ontologías en sistemas de información documental". *El Profesional de la Información* (vol. 20, núm. 5, pàg. 555-563). 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, març-abril). "Topic maps y la indización de recursos electrónicos en la web". *El Profesional de la Información* (vol. 12, núm. 2, pàg. 141-148). <http://www.elprofesionaldelainformacion.com/contenidos/2003/marzo/15.pdf>

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

Roach, M. i altres (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

Enllaços

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

Archives Nationales de France: <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: 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

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

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>

