

La preservació digital des de la perspectiva tècnica

Alice Keefer Riva

P07/09061/02158

Índex

Introducció.....	5
1. La informació digital: els seus elements i les seves amenaces	7
1.1. Naturalesa del recurs digital	7
1.1.1. Formats. Tipologia	9
1.2. Amenaces	12
1.2.1. La degradació dels suports	15
1.2.2. Dades fàcilment alterades	19
1.3. Informació digital i elements per preservar	20
1.3.1. Formes en les quals el recurs pot manifestar-se	21
1.3.2. Elements interns del recurs	22
1.3.3. Propietats essencials del recurs	23
1.3.4. Autenticitat i integritat	25
1.4. Anàlisi de riscos	28
1.5. Registres de formats	29
2. Les estratègies de la preservació.....	31
2.1. La renovació dels suports: el refrescament	31
2.2. Canvis de les dades i del seu entorn: la migració i l'emulació	32
2.2.1. La migració de les dades	33
2.2.2. L'emulació de l'entorn	36
2.3. Altres actuacions	38
3. Els dipòsits digitals i la gestió del material custodiat.....	40
3.1. Dipòsits digitals fiables	41
3.2. <i>Open Archival Information System</i>	44
3.2.1. Components i funcionalitat de l'OAIS	45
3.2.2. Identificació automàtica dels elements tècnics	49
3.3. Avaluació dels dipòsits fiables	50
3.3.1. <i>Audit checklist for the certification of trusted digital</i> <i>repositories</i>	51
3.3.2. Els criteris elaborats per NESTOR	55
3.4. Externalització de serveis de preservació	56
3.5. Els dipòsits institucionals i la preservació	59
Bibliografia.....	61

Introducció

La preservació digital va molt més enllà de l'acte de l'emmagatzemament encara que aquest també sigui important. A diferència de l'acte de només "guardar" la informació digital, la preservació digital està vinculada amb la seva accessibilitat i utilitat per als usuaris del futur i assumeix una gestió activa i, fins i tot, proactiva.

En aquest mòdul s'exploren els principals riscos pel que fa a la longevitat dels documents digitals, les principals estratègies que han sorgit per a detenir la progressiva pèrdua d'aquesta informació i els criteris que les institucions responsables de la preservació dels materials digitals hauran de complir.

1. La informació digital: els seus elements i les seves amenaces

1.1. Naturalesa del recurs digital

Codificació de dades: què és un format? És una codificació estructurada de dades per a la seva gestió i representació mitjançant un programa informàtic. Com en el cas de molts altres conceptes, hi ha definicions variades. La següent, de recent creació i de l'entorn de la preservació digital, va ser elaborada pel Grup de Treball de PREMIS:

"Una estructura preestablerta per a l'organització d'un fitxer digital o una cadena de bits." (*Data dictionary...*, pàg. 4-1)

Alguns formats són estandarditzats, cosa que permet que les dades així codificades siguin recognoscibles per una àmplia gamma de programes. Altres formats són de propietat i requereixen un programari especial per a ser llegits. Dins d'aquest grup, la facilitat de lectura varia molt, entre els que tenen una gran base de clients (per exemple, programes de Microsoft) i d'altres amb una difusió més limitada. El nombre de formats continua creixent, a causa del gran nombre de noves aplicacions i dispositius que entren continuadament al mercat. A més, els documents són cada vegada més complexos, incorporant diferents tipus de formats i funcionalitat.

El format en el qual es crea un recurs depèn del programa d'aplicació. L'elecció d'aquest programa pot respondre a diferents factors, alguns són:

- la plataforma informàtica de la institució a què pertany l'autor,
- la preferència personal de l'autor,
- el propòsit del document i l'estil, qualitat o funcionalitat que pretén oferir,
- i el format necessari per a visualitzar determinats caràcters utilitzats per un col·lectiu; per exemple, l'ús del format LaTeX per a la representació de símbols matemàtics.

Sens dubte, el format dels recursos destinats a ser preservats és un element bàsic per a la presa de qualsevol decisió i pot condicionar:

- la selecció del recurs per a ser preservat,
- el mètode usat per a la seva preservació,
- la facilitat de recuperar i visualitzar el document per a futures generacions,
- el tipus d'ús possible en el futur i
- el cost de la preservació.

Es comentaran les implicacions d'aquestes decisions en els paràgrafs següents.

La selecció del recurs per a ser preservat: tal com es comenta en el mòdul 1, les institucions responsables de la preservació elaboren i apliquen criteris per a triar quins recursos es preservaran. Com és lògic, els criteris estan condicionats per la disponibilitat de recursos –econòmics i humans– per a dur a terme les operacions necessàries.

Entre els criteris per a seleccionar el material digital per a ser preservat hi ha la facilitat per a accedir al document, actualment i en el futur. Per exemple, un criteri de selecció suggerit per Jones i Beagrie en el seu *Handbook* és si el recurs està **actualment** en un format llegible i, fins i tot, si previsiblement el seu format serà llegible **en el futur**. Un factor que facilitarà la llegibilitat en el futur és l'ús de formats estandarditzats –*de iure* o *de facto*– per a la creació de nous recursos. Així, aquests tindran més possibilitats de perdurar per diverses raons:

- les principals estratègies –migració i emulació– seran més fàcils de dur a terme, i
- un nombre més gran d'usuaris causarà més pressió: dit d'una altra manera, es crea un incentiu econòmic perquè empreses del sector s'impliquin en la recerca de solucions per a facilitar la preservació dels recursos en aquest format.

Un format minoritari, per útil i fàcil que en siguin l'ús i la lectura actualment, podrà quedar il·legible i, per tant, inutilitzable. És per això que es pot afirmar que el format del recurs és un factor determinant per a la factibilitat i el cost de la seva preservació.

El mètode usat per a la seva preservació: la selecció d'un determinat format pot obligar que s'adopti una determinada estratègia de preservació. Per exemple, els recursos d'un sol format es presten fàcilment a ser migrats, mentre que aquells que incorporen una varietat de formats i funcions probablement necessitaran l'emulació per a garantir que es preservin en el futur. Però no totes les institucions tindran la capacitat de dur a terme l'emulació i l'alternativa – la contractació de serveis externs– pot representar una barrera econòmica.

La facilitat per a futures generacions de recuperar i visualitzar el document: com s'explica en l'apartat 2 ("Les estratègies de la preservació"), l'emulació pot garantir l'accés al document en el seu estat original. Els programes d'emulació solen elaborar-se segons les necessitats; és el cas del projecte *Domesday* de la BBC: el seu contingut, gravat utilitzant la tecnologia làser al començament dels anys vuitanta, va quedar inaccessible quinze anys més tard.

Què portarà el futur?

Avui en dia és molt difícil preveure com seran les condicions futures, però es pot fer una extrapolació de la situació actual: les biblioteques avui no preserven tot el material tradicional en el seu fons, encara que siguin coneguts tant els problemes com els procediments aplicables. Per què? Pel simple fet que no hi ha prou recursos per a satisfer totes les necessitats. I no sembla haver-hi motiu per a pensar que en el futur les biblioteques estaran més ben dotades econòmicament que en l'actualitat.

El rescat de la informació va ser dut a terme pel ja esmentat projecte CAMILEON (www.si.umich.edu/CAMILEON/domesday/domesday.html), mitjançant l'emulació de l'entorn original.

El cost de les aplicacions es basarà en la complexitat de la feina, amortitzable si n'hi ha una demanda alta. Un format minoritari de baixa demanda requerirà una alta dedicació d'esforç sense l'oportunitat de cobrir les despeses, per la qual cosa resultarà una aplicació costosa que només estarà a l'abast de les institucions de més capacitat econòmica.

El tipus d'ús possible en el futur: un format tancat repercutirà en les futures opcions d'ús. Per exemple, la preservació d'un full de càlcul mitjançant la imatge estàtica (per exemple, GIF) tindria com a conseqüència la pèrdua de la funcionalitat d'aquest tipus de document original. És per això que és imprescindible conèixer la necessitat previsible de la comunitat d'usuaris per a qui es preserva el document. En el cas del full de càlcul, si es coneix el tipus d'ús i el perfil d'usuari es facilita la decisió sobre el format a triar.

El cost de la preservació: sembla evident que un dels "protagonistes" en tots els casos esmentats anteriorment és el factor econòmic. La gran majoria dels reptes es resoldrien amb inversions abundants. El problema és que les institucions culturals responsables tradicionalment de la preservació dels artefactes bibliogràfics no solen estar econòmicament ben dotades. Prova d'això és que, en l'actualitat, la seva tasca de conservació de materials i objectes en suports tradicionals sovint té dificultats per falta de recursos econòmics.

1.1.1. Formats. Tipologia

Quants formats existeixen? És difícil saber-ho del cert, però la National Archives and Records Administration (NARA) dels Estats Units calcula que hi ha uns 16.000 formats presents en els documents que actualment gestiona, provinents de totes les agències de l'Administració pública nord-americana (àmbit federal) (*Jorum preservation watch report*, pàg. 21).

Tampoc no és gaire precisa la tipologia dels formats, ja que entre els experts han sorgit diferències sobre la manera de classificar-los. Actualment hi ha diverses iniciatives en marxa, com el Global Digital Format Registry (GDFR), descrit més endavant en aquesta secció, el propòsit del qual és identificar els formats (i les seves versions i subversions) i mantenir una base de dades amb les seves descripcions per a facilitar-ne la gestió.

A continuació s'ofereixen uns exemples d'esquemes recents en els quals es poden apreciar les diferències en la manera d'agrupar els tipus de formats; per exemple, per gènere, suport o propòsit del document:

- Un estudi recent dut a terme a càrrec de la National Library of Australia (NLA) defineix els grups que s'enumeren a continuació:

- General: gamma àmplia de materials
- Documents d'arxiu, incloent-hi els missatges de correu electrònic
- Materials audiovisuals
- Art digital i literatura electrònica
- Jocs digitals, realitat virtual i programari
- Manuscrits digitals
- Imatges
- Materials miscel·lanis
- Recursos web i d'informació en línia
- Materials digitals en suports físics
- Dades científiques, estadístiques i d'investigació
- Dades geoespaciales i semblants
- Recursos de notícies i d'una altra informació d'actualitat

Segons els autors d'aquest estudi, aquesta llista no és exhaustiva ni definitiva i tampoc no són exclusives les categories. És a dir, hi ha tipus de documents que podrien pertànyer a més d'un grup (*Report to ICABS...*, pàg. 4).

- Les categories següents es van utilitzar en una enquesta duta a terme pel projecte "Rights and Rewards", juntament amb els exemples de formats específics (Bates i altres, 2006, pàg. 9):
 - Document (doc, rtf/rtfd, pdf, xsd, ps)
 - Imatge (jpg/jpeg, gif, png, psd, tif/tff, eps)
 - Àudio (wav, mp3, aac)
 - Vídeo (wmv, avi, rm, mpg)
 - Full de càlcul (xls, xsc)
 - Estadística (SPSS)
 - Diagrama o disseny (Visio, AutoCAD)
 - Bases de dades (SQL, MySQL, Oracle, Access)
 - Presentacions (PowerPoint)
 - Pàgines web
 - Text senzill (txt, XML)
 - Arxius (Zip, Stuffit)
 - Text especialitzat (LaTEX)
- El projecte DELOS va elaborar la tipologia de formats següent [resumida] (*File formats typology...*, pàg. 12):
 - "Semblant" a un document (*document-like*) (aquesta categoria inclou els objectes semblants a un document en paper; és a dir, documents relativament estables que poden contenir text, gràfics, imatges, etc.)
 - Imatges
 - So
 - Imatges en moviment
 - Dades en brut
 - Gràfics estructurats
 - Bases de dades
 - Col·leccions

– Configuració i metadada

La taula següent ofereix les recomanacions del UK Data Archive sobre els formats preferits per aquesta entitat per als documents dipositats, per a facilitar-ne la gestió a llarg termini. Aquesta relació serveix de suggeriment als autors dels formats que afavoriran la preservació a llarg termini, ja que aquest arxiu té com a obligació la preservació de dades resultants dels projectes d'investigació fets al Regne Unit. El seu objectiu és que les dades es puguin consultar, comprovar i reutilitzar en el futur. Com es pot apreciar en el gràfic, els formats es divideixen en tres categories: **preferits** (formats oberts), **acceptables** (formats de propietat de més ús) i **problemàtics**. És lògic imaginar que aquests últims i altres que ni tan sols apareixen en la llista tindran menys possibilitats de ser preservats. La raó de no garantir la seva preservació no guarda relació amb el possible valor de les dades, sinó que es deu senzillament al fet que els recursos econòmics disponibles podran rendir més si es dediquen a fons de més estandardització: és a dir, amb més possibilitats de ser preservats amb èxit i amb un cost menor.

Recomanacions del UK Data Archive per als autors sobre els formats preferits per al material dipositat

UK Data Archive. Formats de dades i programari			
	Format preferit	Format acceptable	Format problemàtic
Dades d'un paquet estadístic	SPSS - arxiu portable (.por) o de sistema (.sav)	STATA; SAS (format xport o dataset versió 7/8 - amb llibreria de formats quan sigui pertinent); text delimitat (separat per tabuladors o per comes)	Format de text d'amplada fixa (no delimitat)
Dades d'una base de dades	Text delimitat amb SQL setup (separat per tabuladors o per comes)	ACCESS, dBase, Filemaker Pro, FoxPro, Paradox, SIR (arxiu d'exportació)	Format de text d'amplada fixa (no delimitat)
Dades d'un full de càlcul	Text delimitat (separat per tabuladors o per comes)	Excel, Lotus, Quattro Pro	
Dades de text	Format de text enriquert (RTF), XML, SGML, text net	Microsoft Word, HTML, WordPerfect, còpia en paper (si aquest és l'únic suport en què estan disponibles les dades de text)	Paquets CAQDAS (p. ex., NUDIST, NVivo, WinMax)
Imatges de mapa de bits	TIFF, PNG	GIF, BMP	Adobe PDF
Imatges vectorials	DXF, SVG	Adobe Illustrator	
Imatges en moviment	MPEG-1, MPEG-4	Quicktime	
So	MP-3	MS Wave (wav)	

UK Data Archive. Formats de dades i programari			
	Format preferit	Format acceptable	Format problemàtic
GIS	Format d'intercanvi/ exportació ESRI ArcGIS (.e00), arxius de forma ESRI ArcGIS	Format d'intercanvi MapInfo (.mif/mid)	MapInfo (.tab)

<http://www.esds.ac.uk/aandp/create/data.asp>

1.2. Amenaces

Hi ha molts tipus de recursos digitals i a causa de la gran varietat de funcionalitats, formats, propòsits, formes de lliurament i usos, és difícil generalitzar sobre la seva naturalesa. Però en un intent de resumir les principals característiques aplicables a la informació digital en general, es destaquen a continuació les dues següents, interrelacionades entre elles:

- la seva dependència d'elements informàtics: la necessitat d'un programa capaç d'interpretar els seus codis perquè es pugui llegir i utilitzar, i
- la seva volatilitat i virtualitat: la facilitat de modificar-la i la seva independència del suport físic.

Aquestes característiques es consideren positives o, almenys, benignes. Però alhora representen problemes reals o potencials per als que pretenguin preservar informació creada en format digital, tal com quedarà exposat en els paràgrafs següents.

Cinc problemes de longevitat

Howard Besser, expert en la digitalització d'imatges i imatges digitals, identifica els cinc factors següents que contribueixen al problema de la longevitat digital:

- **Visualització.** Perquè els usuaris del futur puguin interpretar els documents codificats en anys / generacions / segles anteriors, farà falta el coneixement de la infraestructura original o, almenys, la que sigui necessària per a la seva visualització.
- **Interferències (*scrambling*).** De vegades s'altera el document per raons alienes a la seva preservació. Per exemple, per a estalviar espai, es comprimeix el fitxer; per a assegurar la confidencialitat del sistema, se n'encripten les dades. Aquestes intervencions introdueixen problemes per a la preservació posterior dels recursos, ja que poden comportar pèrdues d'informació (per exemple, la compressió mitjançant mètodes *lossy*) o de l'accés a la informació (per exemple, l'ús de sistemes d'enciptació no desxifrables). És a dir, se sacrifica la usabilitat del document en el futur a causa d'exigències actuals.
- **Interrelacions.** Els recursos digitals es connecten a d'altres, mitjançant enllaços d'hipertext. A més, de vegades el que es visualitza com un sol recurs està compost per elements independents –fitxers amb el seu propi format– que només es reuneixen en el moment de ser consultats (*on-the-fly*). El repte té dos vessants: com preservar els documents enllaçats i com definir les fronteres, és a dir, exactament què cal conservar?
- **Custòdia.** El repte és doble: com s'han d'aplicar les tècniques tradicionals de la preservació per a material intangible i qui se'n responsabilitzarà?
- **Traducció.** La migració de la cadena de bits per a la seva preservació altera el recurs. El repte serà descobrir la manera de salvar no solament el text o imatge, sinó també l'aparença, funcionalitat i "comportament" del document original.

(*Handbook for digital projects...*, pàg. 164-175)

L'obsolescència tecnològica. L'obsolescència tecnològica es refereix a la incapacitat d'utilitzar els elements informàtics –maquinari, programari i suports físics– a causa de l'evolució de la tecnologia. El terme *orfe* (*orphan*) s'utilitza per a referir-se als recursos digitals que, a causa d'aquesta evolució, queden inutilitzables per falta dels elements del seu entorn original. Aquest concepte no significa que els esmentats elements hagin deixat de funcionar, sinó que ja no s'hi pot accedir ni interpretar-los o fer-los funcionar.

Entorn informàtic

La informació digital es crea en un entorn informàtic caracteritzat per una delicada interrelació –de vegades comparat amb un fràgil ecosistema– entre diferents elements que el configuren: un conjunt canviable de maquinari i programari necessari per a crear, emmagatzemar, recuperar i utilitzar les dades digitals. Les dades es creen dins de l'entorn d'un programari d'aplicació que, d'altra banda, arrenca dins d'un sistema operatiu que funciona en uns ordinadors de determinades característiques. Al seu torn, el funcionament dels ordinadors està condicionat pels seus elements integrants, entre els quals hi ha, per exemple, el model del processador i la mida de la memòria.

Aquest compartiment de funcions –una cadena funcional– sempre està en evolució, però no de manera coordinada. Es pot renovar o actualitzar qualsevol de les parts –dades, programes, sistema operatiu, equip físic– individualment, sense tocar els altres elements. Però amb certa freqüència arriba el moment en què comença a fallar-ne un o en què l'equip no pot respondre a les noves exigències adequadament. Per exemple, el processador ja no respon adequadament als requisits de la nova versió del sistema operatiu, la memòria màxima admissible no és suficient perquè les noves aplicacions funcionin en bones condicions, el sistema operatiu no treu tot el profit de les prestacions dels nous programes o no reconeix els nous dispositius disponibles en el mercat.

L'entorn del maquinari. L'evolució de l'equip físic ha experimentat un ritme molt accelerat en l'última dècada. S'ha postulat que el cicle de vida funcional dels equips informàtics és d'entre dos i cinc anys. Molts usuaris han de canviar de sistema, no perquè el seu hagi deixat de funcionar, sinó perquè ja no hi poden baixar les noves aplicacions, per falta de memòria, velocitat, espai en disc, tipus de processador, etc.

A causa de l'evolució tecnològica constant, moltes institucions reconeixen tenir documents digitals en el seu fons, als quals no es pot accedir perquè ja no conserven els ordinadors o altres dispositius necessaris per a llegir-los.

En una possible transició de gran simbolisme, el mateix PC (*personal computer*), que va revolucionar la informàtica fa just vint-i-cinc anys, pot quedar apartat pels dispositius nous –telèfons de tercera generació, sistemes GPS, iPod, Bluetooth, Web 2.0, televisió interactiva. Tot això condueix a l'interrogant: fins quan serà imprescindible l'ordinador, tal com el coneixem?

Exemples de pèrdua d'informació. Un dels primers casos de la pèrdua d'informació per aquest motiu va ser el del primer cens dels Estats Units processat per ordinadors, el del 1960. Uns quants anys més tard, el 1976, l'arxiu nacional nord-americà volia tornar a accedir a aquestes dades d'interès històric. Desafortunadament, aquestes dades es guardaven en cintes només llegibles mitjançant un model de lector que, ja en la dècada dels setanta, era obsolet. Mitjançant un gran esforç i un gran cost van poder recuperar una part d'aquestes dades. L'informe parlamentari sobre aquest cas va declarar anecdò-

PC (*personal computer*)

El va introduir la IBM Personal Computer el 1981. La seva entrada al mercat també va servir per a llançar dos productes que s'incorporaven dins d'aquest model: el microxip Intel i el sistema operatiu de Microsoft que també ha evolucionat: MS-DOS, Windows XP i, últimament, Vista, introduït al començament del 2007.

ticament que en aquella data només existien dos ordinadors capaços de llegir les dades: un al Japó i l'altre al museu Smithsonian, com a relíquia d'una altra època (de tan sols quinze anys anterior!) (*Preserving digital information*, pàg. 3).

D'impacte més recent i proper, molts lectors hauran presenciat la desaparició en els últims anys dels lectors de disquets de 3 ½ polzades, el mateix que va ocórrer fa una dècada amb les disqueteres que llegien i gravaven els disquets *floppy* de 5 ¼ polzades. En un moment donat els fabricants de maquinari deixen d'incloure el dispositiu, per la disminució de demanda o pel seu propi interès a passar a una tecnologia més actualitzada.

L'entorn del programari. Les dades que formen un document digital estan codificades en un format i constitueixen una sèrie de díbits binaris o cadena de bits (*bitstream*). Les cadenes esmentades no tenen cap significat si no hi ha un programari capaç de llegir-les. És per això que, per a descobrir el que està codificat a les cadenes binàries, és necessari el programa que les va crear o un altre que pugui interpretar-ne el significat. Tal com va veure Abby Smith, els «objectes» digitals només existeixen mitjançant la seva recuperació. Sense aquest acte "d'assemblatge" només hi ha una sèrie de bits no desxifrables:

"Quan totes les dades estan registrades com a zeros i uns, no hi ha cap objecte que existeixi essencialment fora de l'acte de recuperació. La petició d'accés crea l'objecte', que és l'acte de recuperació, que provoca una combinació temporal de zeros i uns en una seqüència significativa que pot ser descodificada pel programari i el maquinari." (Smith, 1998)

La complexitat augmenta, ja que el programa d'aplicació necessari per a desxifrar aquests bits es basa en un sistema operatiu que interpreta els seus codis perquè el maquinari pugui funcionar en conseqüència. Si qualsevol d'aquests elements faltés, el programari, igual que les dades, també quedaria orfe. Com podem suposar, l'emmagatzemament dels recursos digitals actuals no aporta cap avantatge si es perd la capacitat de llegir-los en el futur.

Suports físics. Ara com ara, els documents digitals es poden emmagatzemar en suports físics magnètics –discos durs, disquets, cintes DAT, etc.– o òptics – CD-ROM, DVD, etc. Com ja s'ha comentat anteriorment, l'obsolescència tecnològica d'aquests suports està vinculada a l'evolució del maquinari que pot llegir-los i emmagatzemar-los.

El problema pot semblar insignificant i de fàcil solució. Segurament els centres que encara guarden disquets de 5 ¼ polzades trobaran un ordinador equipat amb la disquetera adequada en un altre departament o mitjançant un servei de conversió en una empresa externa. Però sol resultar més difícil del que un es pot imaginar, encara que es tracti d'un suport en actiu fa poc més d'una dècada.

Com es pot llegir un disquet de 5 ¼ " en una universitat espanyola?

Entre el 2005 i el 2006 es van fer un parell d'estudis informals entre bibliotecaris d'universitats espanyoles: els participants havien d'intentar llegir un disquet de 5 ¼ " escollit del seu propi fons. Entre la vintena d'universitats representades, només en una va

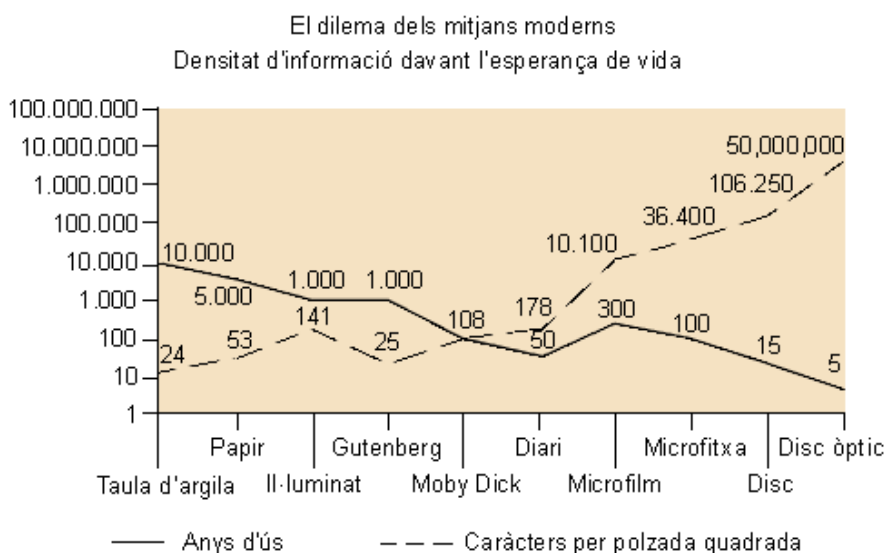
ser possible localitzar un equip adequat per a llegir el disquet en qüestió. En gairebé tots els casos, els bibliotecaris havien de demanar ajuda als informàtics; però molts d'aquests tècnics es mostraven sorpresos per l'interès en material que ells consideraven "antediluviana" o "prehistòric". Segurament no opinarien el mateix sobre els llibres incorporats a la biblioteca durant els mateixos anys que, gràcies al suport de paper, encara es poden utilitzar i el valor dels quals no es qüestiona.

1.2.1. La degradació dels suports

Encara que l'obsolescència tecnològica se sol considerar com el repte més gran a curt termini a causa de la ràpida evolució tecnològica i la despesa que representa la seva actualització, la fragilitat dels suports físics també planteja un problema seriós per a les institucions que custodien recursos en suports digitals –tant els magnètics com els òptics. Afortunadament es tracta d'una amenaça la solució de la qual comporta una bona dosi de sentit comú i és a l'abast de la majoria d'institucions si segueixen les recomanacions. Però no és per això que deixa de ser complexa.

S'entén per la "fragilitat" la inestabilitat intrínseca del suport unida a factors externs com l'ús o les condicions ambientals que poden conduir, tard o d'hora, a la pèrdua de la informació gravada en ell.

El quadre següent va ser publicat en un dels primers informes que van advertir dels problemes de la preservació dels suports digitals (Conway, 1996). L'autor mostra el descens de la longevitat de diversos suports, tradicionals i digitals. Pot semblar irònic que els materials més primitius hagin pogut superar els riscos que hi ha hagut al llarg dels segles, mentre que els més avançats són cada vegada més fràgils per les raons ja exposades.



<http://www.clir.org/pubs/reports/conway2/>

Hi ha molts factors –interns i externs– que afecten la vida d'un suport digital i de les dades gravades en ell. Els **factors interns** estan relacionats amb la composició dels suports, que contenen capes de diferents materials sobre les

quals es graven les dades mitjançant un procés magnètic o òptic. Els suports estan sotmesos a un procés de deteriorament –com si es tractés de l'envelliment d'un ésser viu– que pot provocar errors en la lectura de les dades.

El ritme de la degradació natural d'aquests materials pot avançar-se o agreujar-se a causa de **factors externs**, com, per exemple, les fluctuacions de temperatura i humitat. En els paràgrafs següents s'exposaran els principals factors provinents de l'entorn al costat d'algunes recomanacions sobre el tracte dels recursos físics per a reduir l'efecte danyós. El deteriorament intern –amb efectes com la desmagnetització, la transferència del senyal d'una capa a l'altra i la separació de les diferents capes– també es comenta en l'apartat 2 ("Les estratègies de la preservació").

Els factors nocius poden començar a minar la qualitat del suport des del moment de la seva fabricació, com els següents, que contribueixen a la seva degradació:

- mala qualitat dels materials originals,
- baixa qualitat del procés de fabricació,
- falta de controls ambientals –temperatura, humitat relativa– des del moment de la fabricació del suport, incloent-hi durant el transport i emmagatzemament,
- proximitat a camps magnètics,
- presència de brutícia o pols a l'entorn,
- introducció de fongs i
- mal estat dels equips d'enregistrament i de lectura.

D'aquests factors, alguns són difícilment controlables per la institució: per exemple, la cura dels suports durant el seu transport des de la fàbrica. Però el control sobre altres factors semblaria estar a l'abast de qualsevol institució. És fàcil suposar que una vegada dins de la institució els suports tindran condicions de manteniment i emmagatzemament òptimes. Però la realitat, sovint, és una altra. De vegades, factors institucionals obstaculitzen o impossibiliten que s'adaptin les condicions existents per seguir les recomanacions com ara:

- la rigidesa de criteris de compra que dificulten l'adquisició de material d'alta qualitat,
- la inèrcia o reticència de canviar les maneres de treballar i
- la incapacitat tècnica o econòmica d'efectuar canvis a l'entorn físic, etc.

Malgrat les possibles barreres, què pot fer una institució perquè se segueixin els passos recomanats per a maximitzar la vida dels suports? A continuació es reproduïxen alguns dels suggeriments elaborats per la biblioteca de la Cornell University i dirigits als investigadors i docents que creïn recursos digitals durant el desenvolupament de les seves activitats professionals.

Requeriments per a l'emmagatzemament i manteniment de recursos anterior al seu dipòsit en el dipòsit institucional (*Establishing a central depository*, pàg. 16-17)

Suports

- Emmagatzemar els fitxers mestres en suports digitals [cintes o discos magnètics, discos òptics, etc.] d'alta qualitat fabricats seguint les normes establertes.
- Comprovar els suports periòdicament per a la seva llegibilitat, segons les recomanacions del fabricant.

Còpies de seguretat

- Crear còpies de seguretat dels fitxers mestres i desar-les en un lloc extern segur.

Dispositius d'enregistrament i lectura

- Fer un seguiment dels dispositius d'enregistrament i d'accés per a assegurar que siguin de bona qualitat i estiguin en bon estat (NB: una de les causes més freqüents de la inaccessibilitat d'un document és el mal funcionament d'aquests dispositius o del suport).

Emmagatzemament

- Guardar els suports en un entorn controlat, quant a temperatura i humitat relativa.
- Mantenir nivells consistents, ja que les fluctuacions en la temperatura i la humitat són més danyoses que el no-seguiment dels valors òptims.
- Mantenir-los allunyats de zones amb camps magnètics forts.
- Mantenir net l'entorn de les operacions.
- Minimitzar la manipulació i l'ús de suports d'emmagatzemament magnètics per a reduir-ne el desgast.
- Si es mantenen els suports físics fora de línia, col·locar-los en arxivadors adequats i mantenir-los en posició vertical.

Seguretat

- Controlar l'accés a les zones de processament i emmagatzemament.
- Guardar els suports en una zona apartada, preferiblement accessibles només amb clau.
- Usar sistemes i procediments per a protegir l'autenticitat dels fons i assegurar que no pot fer-s'hi cap alteració, sia a propòsit o accidentalment.

Refrescament

- Els fitxers digitals que durant un ampli període de temps es mantenen a les mans dels seus propis autors després d'incorporar-los al dipòsit institucional, s'haurien de passar a nous suports periòdicament, seguint les recomanacions del fabricant per a determinades condicions mediambientals i fent un seguiment de les tendències cap a tecnologies d'emmagatzemament més eficients. Per exemple, el resfrescament pot ser necessari quan es compren equips d'emmagatzemament nous, quan els nous ja no accepten els suports antics.
- Seguir un procediment de verificació –per exemple, *checksum* o MD5– per a assegurar l'autenticitat i integritat dels suports posterior al procés de refrescament.

Documentació

- Cal explicar per escrit tota acció duta a terme durant el procés de refrescament o d'altres operacions de manteniment que poguessin afectar la integritat dels fitxers.

<http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf>

Com es pot apreciar, són majoritàriament recomanacions de sentit comú i es poden aplicar en qualsevol lloc. Però són essencials: igual que un viatge llarg comença amb el primer pas, assegurar l'accés a la informació digital a llarg termini comença amb les bones condicions de fabricació, emmagatzemament i maneig dels suports físics anterior al fet que el recurs se sotmeti a tècniques de preservació. Els esforços que contribueixen a l'extensió de la vida dels suports i de la llegibilitat de les dades redueixen la freqüència de determinades actuacions de preservació, la qual cosa contribueix a la reducció de costos.

Aquestes recomanacions parlen de les condicions d'emmagatzemament, incloent-hi els controls de temperatura i d'humitat relativa. A continuació s'ofereixen recomanacions elaborades pel Ministeri de les Administracions Públiques (MAP) i també per l'IFLA. Hi ha lleus diferències entre elles i també, en cada cas, entre les condicions recomanades, segons si es tracta de material magnètic o òptic:

**Condicions ambientals recomanades pel
Ministeri de les Administracions Públiques**

- temperatura: 2° - 18° (+/- 1°)
 - suports òptics i suports magnètics
 - humitat relativa:
 - suports òptics: 40% - 55% (+/- 2%)
 - suports magnètics: 40% (+/- 2%)
-

(Ministerio de Administraciones Públicas, 2004, pàg. 30)

Condicions ambientals recomanades per l'IFLA

- temperatura:
 - suports òptics: <20°
 - suports magnètics: 15° +/- 3°
 - humitat relativa:
 - suports òptics: 40%
 - suports magnètics: 30% - 40%
-

(IFLA, 2000, pàg. 78-79)

Què passa si els suports digitals no es guarden en bones condicions? El gràfic següent demostra l'alt grau de degradació, conseqüència de l'increment de temperatura i d'humitat relativa (Jones i Beagrie, 2001, pàg. 130):

Esperança de vida dels suports òptics i magnètics segons les condicions ambientals					
Suport	25 h. r.	30 h. r.	40 h. r.	50 h. r.	50 h. r.
	10 °C	15 °C	20 °C	25 °C	28 °C
D3 cinta mag.	50 anys	25 anys	15 anys	3 anys	1 any
DLT cinta mag.	75 anys	40 anys	15 anys	3 anys	1 anys
CD/DVD	75 anys	40 anys	20 anys	10 anys	2 anys
CD-ROM	30 anys	15 anys	3 anys	9 mesos	3 mesos

Aquest quadre il·lustra la rapidesa amb què pot resultar il·legible la informació digital, sigui quina sigui la qualitat original dels materials de fabricació. La diferència, per exemple, entre una vida de trenta anys o de tres mesos per al mateix suport (és el cas dels CD-ROM) pot representar la diferència entre la preservació, o no, d'un recurs determinat. Un repte per a les institucions que accepten recursos digitals per a la seva custòdia és saber en quines condicions havien estat guardats abans d'ingressar al dipòsit per a determinar quin processament necessiten i amb quin nivell d'urgència.

A més de les condicions de l'entorn, també existeixen recomanacions sobre el maneig dels suports custodiats per a la seva conservació (algunes incloses en les recomanacions ja esmentades de la Cornell University). A continuació es reproduïxen algunes recomanacions de l'IFLA, començant amb algunes de més generals seguides per altres d'específiques, per als diferents tipus de suport (IFLA, 2000, pàg. 76 i 79):

- Generals:
 - Evitar tocar la superfície i, si és necessari, utilitzar guants sense borrisols.
 - Emmagatzemar en àrees netes, sense elements magnètics, pols, fum, etc.
 - Guardar en capsos especials per al suport, en posició vertical.
- Discos òptics:
 - Utilitzar marcadors suaus de punta de feltre sense dissolvents.
 - Netejar amb un paper tissú no abrasiu o, millor, amb una pistola d'aire.
 - Netejar suaument, des del centre cap a les vores del disc.
- Cintes magnètiques:
 - Rebobinar les cintes després de cada ús.
 - Avançar i rebobinar les cintes periòdicament.
 - Girar les cintes anualment.

Els suports magnètics analògics també afectats

Aquesta degradació també afecta altres suports magnètics, encara que el contingut no sigui digital. Un exemple concret: Marc Kirkeby, un alt directiu de l'empresa Sony, parlant de l'arxivament de 600.000 enregistraments sonors analògics en cintes magnètiques, va afirmar el següent: "El desastre potencial no ha de ser *subestimat*. Disposem de cintes del 1949 que sonen perfectament. I altres cintes del 1989 s'han perdut completament. Tot és una qüestió de sort". (Citat a Beauchum)

Recomanació de l'MAP

En canvi, l'MAP recomana que es rebobinin les cintes cada dos anys (Ministerio de Administraciones Públicas, 2004).

1.2.2. Dades fàcilment alterades

A més de la pèrdua de dades a causa dels riscos ja esmentats –l'obsolescència tecnològica i la degradació dels suports–, també hi ha perill de pèrdues espontànies i catastròfiques de la informació preservada. Evidentment, tot el fons bibliogràfic corre el mateix risc davant d'esdeveniments com un incendi o una inundació. Però la pèrdua del fons d'una biblioteca –per traumàtica que sigui– sol ser una pèrdua a escala local i no universal, ja que no se solen eliminar totes les existències de les obres destruïdes: el sistema tradicional d'edició – amb exemplars distribuïts aleatòriament per tot el món – normalment permet que altres còpies dels llibres destruïts es trobin en altres institucions.

Però en el cas dels recursos digitals pot ser que no n'hi hagi més còpies. En una de les nombroses ironies de la nostra època, la facilitat de l'accés als recursos digitals sovint dissimula el fet que són molt poques les còpies mantingudes i, per tant, no existeix la duplicació a gaires llocs, tal com ocorre en el sistema editorial tradicional.

El concepte de tirada en l'edició digital

Una revista científica impresa pot tenir una tirada de 1.500 exemplars que acabaran en sengles col·leccions a diferents racons del món. L'editor se'n quedarà amb uns quants fins que hi deixi d'haver un interès comercial a mantenir el títol. L'eliminació de les existències per part de l'editor o la pèrdua d'una de les col·leccions a causa d'un incendi, no afecten l'existència o la seguretat de les altres col·leccions. En canvi, aquesta mateixa revista en format digital –per exemple, fitxers pdf o html– residirà en un servidor –probablement el de l'editor que segurament mantindrà còpies de seguretat i fins i tot un lloc de rèplica (*mirror*) sota el control d'una altra entitat com, per exemple, una universitat. Però les 1.500 institucions que subscriuen aquest títol no disposen d'una còpia pròpia per a emmagatzemar, ja que mitjançant la seva subscripció els usuaris accedeixen directament a la versió de l'editor o de la rèplica (*mirror*). És a dir, d'un nombre potencial de 1.500 còpies en paper, es passa a la situació gens tranquil·litzant que al món només hi ha mitja dotzena de còpies, incloses les còpies de seguretat, i gairebé totes elles sota el control directe o indirecte de l'editor.

Alguns perills de desaparició sobtada de dades informàtiques són:

- errors humans;
- mal funcionament de sistemes: xarxes, electricitat, telecomunicacions, etc., i
- accions mal intencionades.

És òbvia la necessitat de protegir els suports si es pretén preservar la informació que contenen.

La manera de combatre aquest tipus de pèrdua d'informació és mitjançant estrictes sistemes de seguretat, controls d'accés, còpies de seguretat físicament allunyades del centre, etc.

A més de la facilitat de pèrdua d'informació o de l'alteració de dades que els deixa il·legibles, també hi ha el risc que la informació es perdi a causa de la falta de control de les versions d'un recurs. Aquesta falta de control o la mala identificació de les versions vulnera l'autenticitat i la integritat de les dades. Pot ser el resultat d'una actualització mal documentada o una en la qual no s'han actualitzat les metadades corresponents. De fet, el control ferri de les versions es perfila com un altre gran repte de la gestió a llarg termini dels recursos digitals.

1.3. Informació digital i elements per preservar

A part dels perills ja comentats –obsolescència, deteriorament i pèrdues degudes a la facilitat d'alteració–, la informació digital s'enfronta a un altre repte quant a la prolongació de la seva vida. A diferència dels documents impresos,

hi ha aspectes del document digital que en compliquen el procés de preservació, com ara diferents manifestacions, propietats o altres elements no estandarditzats.

L'objecte que s'ha de conservar

Què interessa preservar: el contingut, el continent o les dues coses? La resposta depèn de la naturalesa del document i de les necessitats de recuperar-lo i utilitzar-lo. Si ens referim a la preservació de documents en paper, és possible que el suport físic sigui més interessant que el text que contingui. En el cas d'un manuscrit medieval, el document físic pot ser tan important com el text. Fins i tot pot resultar insignificant en comparació amb el suport físic. D'altra banda, en un document històric –com, per exemple, un telegrama diplomàtic que informa de l'esclat d'una guerra– la informació té un valor més enllà del suport en el qual es va transmetre.

Per a la majoria dels documents actuals, sol ser prioritari conservar el contingut i no el suport físic –sigui imprès o digital. Una tesina, una memòria d'empresa o uns fulls de càlcul contenen una informació el valor dels quals és completament independent del suport físic, que serveix només per a transmetre o emmagatzemar la informació. La primacia del contingut sobre el contenidor és més patent en el cas de recursos digitals. S'assigna una funció totalment utilitària als suports digitals que s'intercanvien i es retiren sense cap contemplació.

Però els futurs historiadors o arqueòlegs tindran interès a investigar els artefactes digitals d'aquesta generació –discos òptics o cintes magnètiques– de la mateixa manera que els investigadors i col·leccionistes actuals busquen i estudien relíquies quotidianes d'èpoques anteriors (àmfores romanes de fa vint segles, o ampolles de Coca-Cola del començament del segle XX)?

És per això que, tal com s'assenyala en el mòdul 1, cal prendre decisions sobre els aspectes la preservació dels quals sigui necessària i/o factible. Aquesta decisió es basarà en factors com:

- la capacitat tècnica i econòmica de la institució de gestionar l'objecte digital conservat i
- les necessitats d'informació dels usuaris actuals i futurs.

Sense concretar amb precisió l'objectiu –per a què?, per a qui?– de la preservació del recurs, serà difícil escollir la tècnica més adequada per a la seva preservació.

1.3.1. Formes en les quals el recurs pot manifestar-se

A banda de definir exactament quins elements és necessari conservar, també la institució ha de decidir quines formes li interessa preservar i, a més, si és capaç de fer-ho. Se sol parlar dels nivells següents:

- **conceptual o intel·lectual:** allò que l'usuari pot entendre o utilitzar;
- **lògic:** les dades codificades que poden ser interpretades per una aplicació informàtica;
- **binari:** les dades en la seva forma més bàsica, la cadena de bits.

El treball dut a terme pel Grup de Treball PREMIS, per a l'elaboració d'un diccionari de metadades tècniques, va identificar els tres tipus d'"objetes digitals", terme definit per PREMIS com una "unitat discreta d'informació en forma digital":

- **Fitxer:** una seqüència de bytes, ordenada, amb nom i llegible per un sistema operatiu.
- **Cadena de bits:** composta per dades contigües o no contigües dins d'un fitxer, amb propietats significatives comunes en el context de la preservació.
- **Representació:** un objecte digital que representa un objecte intel·lectual. Consisteix en el conjunt dels fitxers digitals emmagatzemats i les metadades estructurals necessàries per a reproduir una còpia completa i acceptable (raonable) de l'objecte intel·lectual.

1.3.2. Elements interns del recurs

A més de les dades codificades en formats llegibles i interpretables per determinats programes, els recursos també poden contenir:

- funcionalitat
- elements multimèdia
- elements dinàmics
- aplicacions interactives
- enllaços d'hipertext

Cada format requerirà uns procediments determinats per a la seva preservació a llarg termini. Com a conseqüència, es complicaran les operacions de preservació per a aquells recursos –cada vegada més nombrosos– que contenen múltiples formats. Aquesta tendència, fàcilment observable, també va ser comentada en un informe recent sobre les activitats de preservació a les biblioteques nacionals: es constata que els recursos actuals són molt més complexos que els que entraven en aquestes biblioteques fa sols uns quants anys (Verheul, 2006). A més dels aspectes tècnics que entren en joc en les decisions sobre els elements que cal preservar i l'estratègia que es vol aplicar, aquestes també estan determinades per altres aspectes, com ara qüestions legals i institucionals que es comenten en els mòduls 3 i 4. Per això, s'ha vist que, com més complex és un recurs, més complexa resultarà la seva gestió i, com a conseqüència, més costosa en serà la preservació.

Alguns elements que formen part del document digital resulten molt difícils, fins i tot de definir; per exemple:

- quina versió es conservaria d'un document actualitzat contínuament?

Manca d'una terminologia estandarditzada

La manca d'una terminologia estandarditzada es fa patent quan els mateixos autors del *Data dictionary* expliquen que la seva definició és diferent de la normalment emprada en l'entorn de biblioteques digitals, en què "objecte digital" es refereix al conjunt de les dades acompanyades pel seu identificador i les seves metadades. [*Data dictionary...*, pàg. 7-3]

- com es representaria un document amb enllaços a documents externs que possiblement no es preservin?

La taula que teniu a continuació, publicada en un estudi britànic sobre els costos de la preservació, projecta la complexitat, i consegüent cost, de la preservació de recursos basats en de la combinació dels tres factors següents: format, titularitat i nivell de control exercit sobre els recursos per la institució responsable de la seva preservació (Granger, Russell i Weinberger, 2000, pàg. 5):

	Més fàcil, econòmic → Més difícil, car
Formats	Pocs → Molts
Drets	Titularitat pròpia → Titularitat aliena
Control	Alt nivell → Baix nivell

1.3.3. Propietats essencials del recurs

Un altre concepte relacionat amb els elements constitutius d'un recurs digital són les anomenades *propietats significatives* (*significant properties*) o elements essencials (*essential elements*), el terme utilitzat per les directrius de la UNESCO. Es tracta del conjunt d'elements considerats essencials per a ser conservats. Aquestes propietats significatives van més enllà del contingut intel·lectual d'un recurs i inclouen altres característiques com, per exemple, la seva funcionalitat i aparença. Inclouen el concepte conegut en anglès com el *look and feel*: aquelles qualitats que representen l'essència del recurs tal com el va desenvolupar el creador i com el van experimentar els usuaris quan es va fer públic; és a dir, l'experiència que tenien en veure'l i tocar-lo/sentir-lo.

Abans de triar l'estratègia de la preservació, la institució encarregada de preservar un recurs digital ha de decidir exactament quins elements és necessari mantenir. Per exemple, la decisió podria ser entre les dues opcions següents:

- preservar el contingut i fins i tot millorar-ne la llegibilitat o utilitat mitjançant les noves tecnologies disponibles en el moment de la seva recuperació;
- mantenir el recurs mitjançant la preservació dels elements que transmetran l'aspecte i comportament (*look and feel*) original, amb les limitacions que pot representar.

Segons la decisió presa per la institució, aquesta podria optar entre les diverses estratègies introduïdes en l'apartat 2; per exemple, la migració del recurs a entorns informàtics actuals o l'emulació, que pretén respectar el funcionament de l'entorn informàtic original. Aquesta decisió serà condicionada, òbviament, pels coneixements dels tècnics i la disponibilitat de recursos institucionals materials i econòmics. Però sobretot és necessari que es tinguin en

compte els interessos del col·lectiu d'usuaris més implicat per a definir els elements de més importància per a ells. Aquells que poden contribuir a aquesta decisió són els mateixos autors, ja que coneixeran quin és l'ús previst per part del públic a qui es dirigeixen.

Una situació semblant al món analògic seria la d'un enregistrament original del cantant de tangos Carlos Gardel. Els aspectes essencials van més enllà de la lletra, la música i la tonalitat de la seva veu: la qualitat de l'enregistrament original permet a l'oient experimentar de certa manera l'efecte original d'aquests "documents" sonors. Una versió actualitzada, amb el soroll de fons eliminat i la veu del cantant retocada, resultaria un producte que possiblement seria útil per als musicòlegs interessats en aquest gènere o els lingüistes que analitzen l'evolució dels dialectes. Però aquesta "millora" hauria tret precisament aquells elements que podrien semblar més autèntics a determinats públics. A més, aquests elements podrien ser considerats essencials pels historiadors que investiguessin la indústria fonogràfica o els biògrafs que escrivissin sobre el mateix cantant. És per això que la decisió sobre quins elements cal preservar ha de basar-se en qui seran els beneficiaris del document preservat: la comunitat designada d'usuaris (*designated user community*), un concepte clau del protocol OAI, descrit en l'apartat 3.2.

Microformes

Un altre exemple del món analògic és el de les microformes, molt utilitzades per les biblioteques a partir dels anys seixanta per a crear facsímils de documents i, d'aquesta manera, protegir els originals. També es feia la microfilmació per a reemplaçar originals que estaven en molt mal estat o per a estalviar espai retirant posteriorment els documents, el suport físic dels quals no es considerava interessant de guardar. Aquest mètode va ser molt ben considerat per ser fidel –o gairebé fidel– a la representació de l'original, captada i reduïda mitjançant un procés fotogràfic. Evidentment no podien ser fidels al 100%, ja que aquestes reproduccions no en conservaven la mida original, encara que, coneixent la reducció exacta, es podia tornar a reproduir sobre paper. Tampoc no se solien conservar els colors. És fàcil imaginar que el color groguenc del paper no es consideraria un element essencial per a preservar, però, en canvi, les obres d'art microfilmades hauran deixat molt a desitjar quant a la seva anàlisi, encara que servissin per a altres usos: per exemple, la cronologia il·lustrada de l'obra d'un artista. És a dir, només determinats col·lectius estan ben servits per les col·leccions microfilmades: continuen essent útils per als que necessiten consultar el contingut intel·lectual però no per a aquells que necessiten analitzar els aspectes ja perduts dels documents originals.

Aquesta diferència entre col·lectius ara s'estén a la capacitat de reconeixement sensorial d'humans a diferència de la dels ordinadors. Pot semblar insignificant la pèrdua d'unes tonalitats d'entre els 24.000.000 colors que poden ser representats per un fitxer TIFF, sobretot si es considera la gran quantitat de gradacions no discernibles per als humans. Però per a una màquina científica, aquesta diferència pot resultar crítica. És per això que, en determinats sectors, els responsables de la preservació hauran de basar les seves decisions en les "necessitats" de futurs usuaris no humans: els equips que seran ocupats cada vegada més en la recerca, recuperació, anàlisi i ús de recursos digitals.

Diversos projectes han analitzat la problemàtica de les propietats significatives, amb l'objectiu d'ajudar els encarregats de la preservació a prendre decisions ben informades sobre quins aspectes cal preservar. Dos dels més coneguts

són CAMILEON (ja esmentat) i DELOS, un projecte europeu que coordina les activitats dels equips que treballen amb temes relacionats amb biblioteques virtuals.

Chris Rusbridge, director del Digital Curation Centre de la Universitat d'Edimburg, fa de "l'advocat del diable" en argumentar que no serà possible guardar totes les propietats essencials dels recursos digitals, per diverses raons (C. Rusbridge, 2006):

- l'enorme quantitat de recursos produïts,
- la immensa gamma de funcionalitats disponibles (una gran part d'elles no utilitzades) i
- la impossibilitat de conèixer quines seran les necessitats dels usuaris del futur.

Rusbridge recomana que les institucions (o la majoria d'elles) es limitin a preservar la cadena de bits i postergar la selecció dels aspectes essencials fins que algú sol·liciti el recurs en el futur. En aquell moment es coneixeran les necessitats dels usuaris i la capacitat de la tecnologia disponible. Aquest autor s'expressa a favor del propòsit plantejat per les biblioteques de la Universitat de Califòrnia de preservar, juntament amb la cadena de bits original, diverses versions dels recursos en diferents formats, cada un conservant diferents elements representatius del recurs en el seu entorn original. Aquestes versions les anomenen *dessecades (desiccated)*, a causa de la inevitable pèrdua d'elements i prestacions en cada una d'elles, com, per exemple, animació, enllaços, colors, gràfics intercalats amb text (*Overview...*, pàg. 5).

1.3.4. Autenticitat i integritat

El canvi de la naturalesa dels recursos que cal preservar en l'era digital ha alterat els conceptes d'autenticitat i integritat, i també la manera d'assegurar que els documents es mantinguin autèntics i íntegres. Alhora la facilitat amb què es poden alterar les dades fa que sigui molt urgent trobar un mecanisme que comprovi que el document és el que pretén ser i que les dades no han estat canviades, accidentalment o amb intencionalitat.

A més del repte de redefinir aquests conceptes en el nou entorn i d'elaborar noves formes de comprovar l'estat del recurs, l'anàlisi d'aquest assumpte es complica més per la falta de definicions homologades. L'ambigüïtat resultant es deu en part al fet que el significat d'aquests termes varia segons el sector professional. En paraules d'Abby Smith:

"El terme [*autenticitat*] té un significat específic per a un arxiver i també específic però una mica diferent per a un bibliotecari de llibre antic/rar; alhora hi ha criteris diferents per a avaluar l'autenticitat per a materials publicats o no publicats." (*Authenticity in a digital environment*)

Les directrius de la UNESCO defineixen els conceptes d'autenticitat i integritat de la manera següent (*Directrices...*, pàg. 115, 165-166):

L'**autenticitat** d'un objecte resulta alhora de la confiança en la seva *identitat* (és el que diu ser i no es confon amb cap altre) i en la seva *integritat* (no ha experimentat cap modificació que en canvi el significat) (pàg. 115).

Autenticitat. Garantia del caràcter genuí i fidedigne de certs materials digitals, és a dir que són el que s'afirma d'ells, sia objecte original o còpia conforme i fiable d'un original, feta mitjançant processos perfectament documentats (pàg. 165).

Integritat d'objectes digitals. Estat dels objectes que es troben complets i que no han patit corrupció o cap alteració no autoritzada ni documentada (pàg. 166).

Autenticitat i integritat segons diferents fonts

Diferents fonts descriuen l'autenticitat i la integritat com dos elements relacionats però diferents, mentre d'altres els veuen formant una relació jeràrquica. Però fins i tot hi ha una divergència de definicions en aquest últim cas: alguns consideren l'autenticitat com un component de la integritat i altres, al revés, inclouen la integritat dins del concepte de l'autenticitat. Els exemples següents són definicions exposades per diferents experts:

Bearman i Trant: Els judicis sobre l'**autenticitat** es basen en valoracions sobre els orígens, la completesa i la **integritat** interna d'un document... L'**autenticitat** (o la garantia del contingut) fa referència a allò que el recurs pretén ser i a la manera en què es va crear i, per tant, està basada en mètodes d'identificació i verificació del recurs. (*Authenticity of Digital Resources*)

Coalition for Networked Information (Coalició per a la Informació a través de Xarxes): **Autenticitat** ...pot significar ser original, però també ser fidel a un original; pot significar una procedència no corrompuda, però també una procedència clara i coneguda, sigui "corrupta" o no. Rere cada definició d'**autenticitat** hi ha assumpcions sobre el significat i la significació del contingut, la fixesa, la coherència de les referències, la procedència i el context. (*Defining Authenticity...*)

Conway: "L'**autenticitat** o veracitat del contingut de la informació d'un article... És la base de la **integritat** intel·lectual." (*Handbook for digital projects...*)

Cullen: "Un objecte **autèntic** és aquell la **integritat** del qual resta intacta—un objecte del qual es pot demostrar o acceptar que és allò que el seu propietari diu que és." (*Authenticity in a digital environment*)

Hirtle: A l'hora d'establir la **integritat** dels registres, la custòdia ininterrompuda és tan important com la completesa dels documents. (*Authenticity in a digital environment*)

International Council of Archives (Consell Internacional d'Arxius): **Autenticitat** = un document és el que pretén ser. Fiabilitat = un document és fidedigne. (*Authenticity of Electronic Records*)

Lynch: **Integritat**... volem dir que [el document] no s'ha corromput amb el pas del temps o durant la seva circulació; és a dir, que tenim a les mans el mateix grup de seqüències de bits que va néixer en el moment de crear l'objecte. (*Authenticity in a digital environment*)

Rothenberg: "*Autenticitat* ... se suposa que inclou qüestions d'**integritat**, completesa, exactitud, validesa, fidelitat a un original, significació i adaptació a la finalitat pretesa." (*Authenticity in a digital environment*)

Malgrat la varietat d'interpretacions existents o la definició preferida per diferents institucions o col·lectius, el problema de fons és evident. Tant la integritat d'un document —és a dir, si es troba complet o íntegre— com l'autenticitat —si

el document realment és el que afirma ser— són difícils de garantir en l'entorn digital. Si el recurs digital no té una versió física per a dur a terme una comprovació (per exemple, la foto o el manuscrit original), cal fiar-se de la descripció disponible de les seves característiques per a determinar si és complet, o si representa una alteració de la versió original. Aquesta descripció també ha d'incloure dades sobre qualsevol modificació que s'hagi fet: per exemple, el canvi de format com a resultat de la migració. Aquesta informació es troba en les metadades, que per això tenen una importància de primer ordre en la preservació digital.

La dificultat de garantir l'autenticitat del cert explica la importància que té el paper de "fiabilitat" (*trust*) en la preservació: els usuaris del futur hauran de confiar en la institució que custodia els recursos, en els mètodes emprats en la custòdia i en el recurs recuperat. És per això que la solució proposada és la creació i certificació de dipòsits digitals *fiabls* (*trusted digital repositories*), descrits en l'apartat 3 ("Els dipòsits digitals i la gestió del material custodiat").

Com es pot actuar per a assegurar la integritat i l'autenticitat? En el cas de la integritat, la informació pot perdre's de diverses maneres, com, per exemple, accidents, avaries o actuacions humanes danyoses, entre d'altres, la negligència o els actes vandàlics. És per això que la protecció es basa sobretot en:

- el manteniment dels equips d'enregistrament, emmagatzemament i lectura en bones condicions,
- l'adequada formació del personal,
- la protecció de la infraestructura i els recursos.

Algunes aplicacions informàtiques poden comprovar la integritat del fitxer i les còpies de seguretat poden assegurar la restitució de recursos que hagin patit danys o pèrdues sempre que aquestes còpies es facin sistemàticament i es mantinguin allunyades dels originals. Per a comprovar l'autenticitat del recurs es requereixen sobretot un bon control de versions i un mecanisme per a detectar i documentar alteracions. És evident, doncs, que per a garantir l'autenticitat i la integritat dels recursos, les tècniques de descripció i seguiment són tan importants com els procediments informàtics.

Adam Rusbridge exposa les tres etapes en què les dades corren més perill de resultar alterat. És per això que són els principals punts en els quals s'haurà de poder demostrar la fiabilitat de les dades (A. Rusbridge, 2003, pàg. 16-19):

- **Abans d'introduir-se al dipòsit:** caldrà esbrinar si el recurs, a punt de ser incorporat al fons preservat, és el que s'afirma que és. Per exemple, si un recurs es presenta com a elaborat amb el format WordPerfect 3.5 per a Macintosh, el dipòsit necessita tenir un mecanisme que ho verifiqui [com el sistema JHOVE, descrit en l'apartat 3, "Els dipòsits digitals i la gestió

del material custodiat"]. A partir d'establir la seva autenticitat, es triarà la millor manera de mantenir-lo accessible i utilitzable.

- **Durant la custòdia:** per a mantenir l'autenticitat del recurs una vegada dipositat, es requereixen bones mesures de seguretat i de manteniment. Per exemple, hi pot haver una pèrdua de funcionalitat durant la migració de formats o alteracions degudes a una apagada o un virus informàtic. És per això que les eines de protecció de què el dipòsit ha de disposar inclouen des d'aplicacions informàtiques fins a polítiques de gestió.
- **En el moment de recuperar-se:** els usuaris hauran de fiar-se del recurs que els lliuri el dipòsit, en qualsevol moment del futur. Podria ser una còpia idèntica a la versió emmagatzemada, però també és possible que es generi sobre la demanda. En aquest cas seria una versió feta *ex professo* a partir de la cadena de bits original o d'un format intermedi utilitzat pel dipòsit per a facilitar la gestió dels recursos dipositats.

El contingut, aspecte i funcionalitat de la versió lliurada a l'usuari es formarien amb aquelles propietats essencials definides en el moment de la seva incorporació en el dipòsit. És a dir, pot no ser una rèplica exacta del que es va crear originalment. Però tot i així, podria considerar-se autèntic si estan documentats els canvis a què ha estat sotmès el recurs. Per exemple, un informe creat originalment en WordPerfect pot ser lliurat a l'usuari en la versió actual de PDF. Haurà perdut alguns elements del seu entorn original i, alhora, se n'hauran incorporat d'altres, per exemple, les prestacions de PDF que no oferia WordPerfect. El que és important és que sigui, efectivament, el que pretén ser i que l'usuari sàpiga en què difereix la versió rebuda de l'original.

1.4. Anàlisi de riscos

Una activitat recomanada per a identificar les amenaces en cada centre –donat el tipus de fons i determinades condicions del seu entorn– és fer una anàlisi de riscos (*risk analysis*). Amb el reconeixement que no es podran evitar tots els riscos, l'anàlisi de les possibles amenaces permet preveure les més probables i prendre decisions sobre la millor forma de protegir-se contra elles mitjançant l'aplicació dels mètodes més adequats; és a dir, una gestió dels riscos (*risk management*). Les decisions es basaran en factors com els següents, presos del model d'anàlisi d'amenaces publicat en les directrius de la UNESCO (*Directrices para la preservación...*, pàg. 123-124):

- el tipus de dany
- la probabilitat que ocorre
- la rapidesa de l'aparició
- les repercussions
- les opcions de prevenció

A tall d'exemple, es resumeix a continuació l'anàlisi de dues de les amenaces identificades:

Amenaça	Tipus de dany	Probabilitat	Rapidesa	Repercussió	Opcions
Les avaries del suport	Danyen la integritat de les dades.	Tenen una alta probabilitat d'ocórrer.	Apareixen gradualment.	Tenen repercussions greus, ja que les dades poden resultar il·legibles i irrecuperables.	Es poden prevenir mitjançant l'ús de productes d'alta qualitat, suports més estables, controls freqüents, transferència de dades.
Les catàstrofes naturals	Poden danyar la integritat de les dades, la identitat dels fitxers i els equips.	Tenen una alta probabilitat d'ocórrer a llarg termini.	Apareixen sobtadament, però hi pot haver un avís previ.	Tenen repercussions que poden ser mínimes i localitzades o molt greus.	Es poden prevenir mitjançant magatzems ben ubicats, còpies de seguretat externes, etc.

1.5. Registres de formats

Un altre dels reptes relacionats amb els formats és com preservar la informació que explicarà a futures generacions com eren (i com era cada una de les seves versions i subversions). Quan un programari canvia de versió, el resultat són diferències de format, més o menys, entre documents creats en una o una altra versió. A més, el format també varia segons el sistema operatiu en el qual ha de funcionar el programari. La quantitat de formats, amb les seves múltiples variacions, dificulta la lectura de la cadena de bits dels recursos custodiats, sobretot després del pas de diverses generacions informàtiques. Fins i tot els formats oberts i estandarditzats van modificant-se com a resposta a l'evolució tecnològica i a les exigències dels que els utilitzin.

Encara que perduri alguna forma d'un determinat format, és possible que les versions més antigues deixin de ser reconegudes pels futurs filtres de conversió. Però les institucions que conservin recursos codificats en aquestes versions necessitaran disposar d'informació específica sobre elles.

Una iniciativa molt prometedora per a recopilar aquesta informació –i així evitar que cada dipòsit ho hagi de fer– és el projecte **Global Digital Format Registry** (GDFR) (Abrams, 2005). El seu propòsit és recollir, emmagatzemar i subministrar als dipòsits digitals informació completa i fiable sobre tots els formats gestionats per ells. En aquest projecte de dos anys, subvencionat per l'Andrew W. Mellon Foundation i coordinat per la biblioteca de la Universitat de Harvard, es desenvolupa el registre definitiu sobre unes bases definides durant diverses reunions celebrades abans de l'inici del projecte, en les quals participaven experts del món bibliotecari. Entre les biblioteques i arxius nacionals que van col·laborar en la fase de planificació, hi figuren les següents institucions: Bibliothèque Nationale de France, British Library, Library of Congress, National Archives and Records Administration, National Library of Canada, The National Archives (Regne Unit).

Tal com expliquen els responsables del GDFR, la informació sobre els formats que recollirà el registre és imprescindible per a poder interpretar correctament el contingut dels recursos digitals, ja que sense ella, "un objecte digital és senzillament una col·lecció de bits no diferenciats" (Abrams i Flecker, 2005, pàg. 1).

Aquesta informació –coneguda com a *informació de representació*– ha de figurar en les metadades dels recursos custodiats i permet que puguin ser llegibles i utilitzables en el moment de la seva recuperació.

Cada entrada del registre conté informació descriptiva sobre un format determinat, les seves propietats i també els sistemes o programes que poden reconèixer, manipular o crear recursos en aquest format. Aquesta informació també inclou dades més aviat administratives com la història o procedència del format (Abrams, 2005).

2. Les estratègies de la preservació

En la secció anterior s'han presentat els riscos a què estan exposats els recursos digitals. En aquesta es repassen les accions pal·liatives per a contrarestar alguns dels perills descrits. L'estat de la qüestió sobre com s'ha de plantejar la preservació de documents digitals revela diverses tendències, encara que cap d'elles no resolgui totalment el problema. A continuació s'examinen les principals opcions amb els seus avantatges i inconvenients. Aquestes no se solen aplicar en exclusiva, ja que els centres empraran una varietat d'accions, atenent factors ja introduïts anteriorment, com ara:

- el format dels recursos,
- els usos previstos per a ells, i
- la capacitat tècnica institucional disponible.

A més, cal recordar que les solucions definitives es desenvoluparan paral·lelament amb:

- l'evolució de la tecnologia, que oferirà més alternatives, encara desconegudes;
- la coordinació de totes les entitats implicades (*stakeholders*) que aportarà una millor utilització dels recursos existents, i
- el desenvolupament i l'aplicació de normes per a assenyalar el camí.

En aquesta secció, abans de parlar dels mètodes emprats per a evitar l'obsolescència tecnològica i assegurar que els recursos siguin accessibles i utilitzables en el futur, es comenta una altra acció necessària per a preservar la vida dels recursos digitals: la renovació o refrescament dels suports per a evitar que la informació es perdi perquè es deterioreni.

2.1. La renovació dels suports: el refrescament

La transferència de les dades d'un suport a un altre de nou –coneguda com a *refrescament*– serveix per a reduir el risc de la pèrdua de dades deguda al deteriorament que pateix tot suport, tal com s'ha explicat en la secció anterior. En realitat no s'hauria de considerar una estratègia, ja que no és una opció a elegir sinó una acció imprescindible, igual que les recomanades per al maneig físic dels suports. El refrescament s'inclou en aquesta secció perquè moltes fonts el tracten al costat de les principals estratègies. A més, necessita una programació rigorosa, semblant a la que s'haurà d'implementar per als altres processos de preservació.

La transferència de les dades d'un suport a un altre s'efectua sense causar cap canvi en el programari o el format; és a dir, es refresca l'entorn. El contingut es renova sense alterar el contingut. A vegades, el refrescament coincideix amb un canvi generacional del suport, com quan es passa informació de cintes o discos originals a models actuals.

La freqüència amb què es refresca varia segons la política institucional i el tipus de suport, però sempre s'hauria de fer dins del termini recomanat pel fabricant. Els factors institucionals que determinaran el calendari exacte de les actualitzacions inclouen, per exemple, el tipus d'ús i les condicions d'emmagatzemament. Com queda patent en el quadre sobre l'expectativa de vida dels suports digitals, elaborat per Jones i Beagrie i exposat en la secció anterior, la degradació pot ocórrer a un ritme accelerat si les condicions ambientals no són les adequades.

En resum, el refrescament no ofereix cap solució per a protegir les dades del risc de l'obsolescència tecnològica, ja que romanen en el seu format original. Però és un pas –repetible i imprescindible– en el procés de la preservació. No implica una gran inversió en equips ni un alt grau de coneixements tècnics per part del personal. En canvi, requereix una gestió activa: la programació del refrescament segons les indicacions dels fabricants i les millors pràctiques identificades per la professió, el seguiment del calendari establert i la descripció mitjançant les metadades tècniques de les accions realitzades. També és important que s'observi l'evolució del mercat per a triar sempre els suports que ofereixin una vida més llarga.

2.2. Canvis de les dades i del seu entorn: la migració i l'emulació

Com ja s'ha comentat, hi ha dues principals amenaces per als recursos digitals: la fragilitat dels suports i l'obsolescència de l'entorn informàtic original. En la secció anterior s'ha demostrat com l'acte de refrescar el suport és necessari per a combatre la pèrdua d'informació deguda al deteriorament de les unitats d'emmagatzemament, tant magnètiques com òptiques.

En aquesta secció l'enfocament s'orienta cap a les mesures que prenen les institucions per a intentar prevenir la pèrdua de la informació a causa del desfament de l'entorn digital –maquinari, programari, suports i formats– i evitar així que el recurs quedi inaccessible o, en altres paraules, "orfe"¹.

Avui dia se solen emprar dues estratègies principals: la migració i l'emulació. Cada una té avantatges i inconvenients i pot adaptar-se amb més o menys èxit a diferents tipus de fons digitals. Actualment, totes dues són emprades tant per centres que preserven els seus propis documents com per altres que assumeixen la responsabilitat de preservar recursos aliens. Aquests últims inclouen les biblioteques nacionals i les empreses de serveis contractades per institucions que manquen d'una infraestructura tècnica adequada.

Còpies de seguretat

De fet, el procés de transferència de dades d'un suport a un altre és semblant a l'utilitzat per a la creació periòdica de còpies de seguretat i, per això, sovint es confonen els conceptes. A diferència del refrescament, les còpies de seguretat no estan pensades per a la preservació, sinó per a la recuperació en cas de pèrdua de dades. Els suports on es graven les còpies de seguretat solen reutilitzar-se per a l'enregistrament de noves actualitzacions. És a dir, la **informació** s'actualitza en comptes dels **suports**, com en el cas del refrescament.

⁽¹⁾També s'utilitza la paraula *orfe* (*orphan*) per a referir-se a documents sense el titular localitzable. Vegeu el mòdul 4 sobre els aspectes legals de la preservació digital.

A més d'aquests dos mètodes, també hi ha una sèrie de pràctiques que, si no arriben a ser considerades per tots com a estratègies, sí que representen una manera d'actuar, encara que sigui per defecte. Aquests inclouen:

- el manteniment dels museus informàtics,
- la reproducció en paper o en microformes, i
- l'arqueologia digital.

2.2.1. La migració de les dades

L'estratègia de migració té com a funció principal la conversió del **document** creat en un determinat entorn i codificat en un determinat format a un altre format perquè funcioni en una nova plataforma informàtica: una de més actual o més estandarditzada. És a dir, el document s'adapta als nous entorns a diferència del que ocorre en l'altra estratègia principal –l'emulació (comentada en la pròxima secció)–, en la qual l'**entorn** informàtic actual intenta reproduir la funcionalitat de l'original, sense que s'efectuïn canvis en el mateix document.

La definició de migració següent és l'oferta pel *tutorial* de la Universitat de Cornell titulat *Digital preservation management*:

"[...] un conjunt de tasques organitzades i dissenyades per a aconseguir la transferència periòdica de materials digitals des d'una configuració de maquinari i programari a una altra, o des d'una generació de tecnologia informàtica a una generació posterior."

(Kenney i altres, 2003)

La migració és l'estratègia més emprada avui dia, en part perquè és un procés informàtic molt experimentat. Tanmateix, comporta el risc que es perdin elements especials –com, per exemple, els codis que controlen el format o la funcionalitat de l'entorn original– que no sempre es poden preservar durant la conversió.

La migració està especialment indicada per a grans lots de recursos d'un sol format. La possibilitat de programar per endavant les migracions per a dur-les a terme en el moment oportú contribuirà a mantenir els costos més baixos. També es preveuen estalvis si s'arriba a reduir la freqüència amb què sigui necessari migrar els formats. Això s'aconseguirà mitjançant l'adopció de formats normalitzats per a la creació de recursos en comptes de l'ús de formats de propietat per més actualitzats que siguin.

Si la migració es duu a terme en el moment oportú, és relativament senzilla. Per a això es recomana un seguiment constant dels avenços en la tecnologia, mitjançant un "observatori tecnològic" (*technology watch*) que permetrà preveure els canvis esdevenidors i així prendre les decisions adequades sobre la selecció del nou entorn i programar a temps la migració.

Migracions de dades bibliogràfiques

En el món de les biblioteques, la migració pot ocórrer quan una biblioteca canvia d'un sistema de gestió bibliogràfica a un altre. Si les dades bibliogràfiques són estàndards –per exemple, si segueixen el format MARC– poden ser migrades sense problema. Però la transferència d'altres dades no normalitzades –les d'usuaris, préstecs o proveïdors– sol ser més problemàtica i de vegades no es pot dur a terme, deixant com a solucions una conversió costosa o la reintroducció manual de les dades.

Què passa si no es fa la migració de les dades en el moment oportú? En organitzacions que han dut a terme migracions de programari, l'experiència demostra que sovint els documents que no s'utilitzen quan es fa la migració romanen emmagatzemats en el seu format original i corren el risc de convertir-se en orfes irrecuperables quan el seu entorn quedi obsolet. És important reconèixer que hi sol haver un període de durada variable en el qual la transferència pot dur-se a terme fàcilment.

La migració no pot ser programada tan sistemàticament com el refrescament dels suports, el calendari dels quals respon a les recomanacions dels fabricants i a l'expectativa de vida prevista. En canvi, l'evolució dels formats respon a factors externs, impulsats per la indústria informàtica segons les pressions i oportunitats del mercat, i per les agències internacionals que treballen en la normalització de formats per a evitar que la diversitat de formats inhibeixi l'avenç de la ciència i, també, del mercat.

De fet, es tracta d'una de les nombroses contradiccions que marquen la diferència de tractament entre un document digital i un altre en suport tradicional. En el cas d'aquests últims, la falta d'ús es considera un element a favor de la conservació. Seria el cas d'un diari manuscrit que ha quedat intacte dins d'un armari durant l'últim segle, protegit en un microambient tancat. Probablement el seu estat serà millor que el d'un altre document que estigués deteriorat pels danys produïts per cent anys d'ús físic i efectes mediambientals nocius. Aquest tipus de preservació "passiva" o abandonament benigne és positiu per a la majoria dels documents tradicionals, però perjudicial per als digitals, que poden quedar inaccessibles en pocs anys.

Hi ha diferents varietats de migració, com les següents:

Migració sobre la demanda: es tracta de separar el codi binari del seu entorn i suport original en el moment en què es dipositi el recurs. Juntament amb aquesta cadena de bits, es guarden les metadades tècniques que especifiquen com s'ha de migrar a altres entorns. D'aquesta manera el centre evita haver de fer migracions periòdiques de tota la col·lecció per a poder concentrar els esforços i recursos en els documents amb una demanda constatada. Encara que mitjançant aquest mètode s'estalviïn els esforços de les migracions periòdiques, alhora es requereix més dedicació en la preparació i l'actualització de les metadades tècniques que s'han de modificar paral·lelament a l'evolució de la tecnologia (*Cedars guide to digital preservation...*, 2002).

Migració a un format estàndard: es refereix a la migració del document original a un format neutral o estandarditzat, amb una durada prevista més gran. Per exemple, molts centres passen els documents textuais de diferents formats a XML, PDF o altres formats oberts. Per aquesta raó i per disposar, com a conseqüència, d'un nombre més gran d'usuaris, es preveu que aquests formats tindran una vida més llarga que un format de propietat.

La programació de la migració

L'Atmospheric Radiation Monitoring Center del Laboratori Nacional d'Oakridge (Tennessee, Estats Units) calculava que era necessari migrar a noves tecnologies cada quatre o cinc anys, i que cada migració durava entre sis i dotze mesos (Hodge, 2000).

En una mesura semblant, alguns dipòsits només admeten determinats formats. D'aquesta manera, els dipòsits poden centrar els seus esforços i recursos en la gestió del fons digital en comptes de malgastar-los processant uns formats minoritaris. Algunes institucions incentiven o fins i tot obliguen els autors a utilitzar formats estàndard per a crear documents. En canvi, d'altres assumeixen la responsabilitat de convertir el fitxer a un format estandaritzat en el moment d'ingressar el document.

Migració successiva o conversió en sèrie: mitjançant aquesta tècnica (coneguda en anglès com a *serial conversion* i també *backward compatibility*) s'actualitzen els documents amb cada canvi incremental del programari, tant si és necessari com si no ho és per a l'operació interna. És obvi que quan s'introdueixen noves versions de programari és justament quan hi ha més motiu per part dels fabricants per a facilitar-ne la conversió. D'aquesta manera, no corren el risc de perdre l'oportunitat, quan el format estigui a punt de desaparèixer i hi hagi menys facilitats per a fer la migració.

La migració té diversos inconvenients, entre els més importants destaquen els següents:

- No es pot evitar l'alteració del document original. Aquesta alteració pot afectar-ne l'estructura, el contingut, l'aparença i la funcionalitat més o menys. Per la inevitabilitat d'algunes alteracions, els processos seguits durant la migració han de ser documentats mitjançant les metadades tècniques, igual que qualsevol efecte que hagin produït en el document. D'aquesta manera els futurs usuaris sabran quines diferències hi ha entre el document original i la versió recuperada.
- No és gaire adequat per als objectes digitals que contenen una varietat de formats, ja que cada un necessitarà rebre un tractament diferent i així serà molt difícil l'automatització del procés. Actualment, la catalogació tradicional no capta amb exactitud la presència dels diferents elements incorporats al document i dificulta encara més la migració de cada format en el moment oportú.

Elecció de formats

Coneixent les necessitats es pot escollir entre estratègies (migració, emulació, etc.) o entre els formats als quals es convertirà el document original: per exemple, un document textual que podria migrar-se a un altre format de text o com a imatge (per exemple, JPEG). Aquest últim captura el contingut i el format amb exactitud, però no tindrà la capacitat de fer cerques sobre el text, prestació que podria ser essencial per a determinats col·lectius com, per exemple, lingüistes, que estudien les concurrències de paraules. La decisió sobre si cal mantenir aquest document en format de text, amb la probable pèrdua de format, o convertir-lo en un altre d'imatge hauria de prendre's pensant en el probable ús futur que se'n faci.

Es resumeixen a continuació els principals avantatges i inconvenients de la migració:

Avantatges

- És una operació molt experimentada.
- No es requereixen coneixements tècnics gaire especialitzats.
- Es pot automatitzar una part dels processos.
- La tendència cap a l'estandardització de programari i formats facilitarà la migració.
- Converteix el document a un format compatible amb sistemes actuals.

Inconvenients

- És difícil programar la migració per a objectes digitals complexos.
- Altera el document amb el risc de la pèrdua d'elements importants.
- Ha de repetir-se periòdicament al llarg de la vida del recurs.

2.2.2. L'emulació de l'entorn

Els partidaris de l'emulació creuen que aquesta estratègia ofereix la millor solució per a preservar recursos digitals més complexos, ja que és l'única que garanteix la recuperació del document original, sense alteracions com les inevitables que ocorren en la migració. L'objectiu és que els sistemes informàtics del futur tinguin la capacitat de recuperar les dades originals, com si es tractés del programari original. Així, el programa emulador permetrà que l'usuari del futur vegi i experimenti el recurs tal com era quan es va crear. Com s'ha afirmat en la secció anterior, en comptes de pretendre convertir **el document** original perquè funcioni en sistemes nous, l'emulació converteix **l'entorn** perquè pugui llegir el document en el seu estat original.

En el seu informe *Avoiding technological quicksand* ('Com es poden evitar els sorramolls tecnològics'), Jeff Rothenberg, el primer proposador d'aquesta estratègia, argumentava que l'emulació era l'única manera capaç –almenys en teoria– de garantir la conservació de les dades juntament amb els seus elements essencials, pel que fa tant a la seva aparença com a la seva funcionalitat –és a dir, el seu *look and feel*– "de manera coherent, automàtica i homologable per a diferents tipus de documents" (Rothenberg, 1999).

A causa dels avantatges d'aquesta estratègia, el projecte europeu NEDLIB va triar l'emulació per a desenvolupar el prototip d'una biblioteca europea virtual de dipòsit de material digital. El desenvolupament de l'emulador es va encarregar al mateix Rothenberg. Els resultats d'aquest projecte van contribuir al disseny i implantació del sistema actual emprat per la biblioteca nacional holandesa.

A diferència de la migració, l'emulació no és tan experimentada i requereix uns coneixements informàtics més avançats, per a entendre la funcionalitat del sistema original i programar els sistemes actuals perquè la reproduueixin. Segons les conclusions de NEDLIB, encara faltaven experiències reals per a comprovar el funcionament de l'emulació a la pràctica. El 2000 Rothenberg resumia el que quedava per fer per a desenvolupar un sistema operacional (Rothenberg, 2000, pàg. 69):

- Desenvolupar tècniques d'especificacions que puguin descriure tots els atributs pertinents de les plataformes informàtiques, amb prou detall perquè puguin ser emulades en sistemes del futur.
- Desenvolupar tècniques per a l'allotjament, amb un mínim d'esforç, d'aquests emuladors a qualsevol plataforma del futur.
- Continuar retocant les metadades necessàries per a descriure els documents digitals i per a enllaçar-los amb el conjunt de programari i maquinari emulat que els permetrà accedir-hi en el futur.
- Assegurar que les descripcions podran mantenir-se en format llegible a llarg termini.
- Continuar identificant i retocant els criteris d'autenticitat i les proves que validen els recursos digitals de diversos tipus, i així oferir un mecanisme per a avaluar l'èxit de la preservació basada en l'emulació.

A continuació es resumeixen els principals avantatges i inconvenients de l'emulació:

Avantatges

- Reprodueix el document original, incloent-hi el *look and feel*.
- No requereix un seguiment continuat del format del recurs.
- Ofereix una solució per a objectes digitals complexos, ja que no s'ha de controlar cada tipus de format i funcionalitat que continguin.

Inconvenients

- Hi ha poques experiències reals.
- S'hauran d'actualitzar els emuladors amb cada canvi significatiu de sistemes.
- La programació dels emuladors és complexa i requereix coneixements especialitzats, la qual cosa representa un cost més elevat.
- Depèn de la predisposició dels informàtics en el futur.

Programes d'emulació

Un estudi recent de l'estat de la preservació digital a les biblioteques nacionals va afirmar que dels programes d'emulació existents, molts serveixen per a motius "nostàlgics" com, per exemple, recrear les plataformes informàtiques originals dels jocs informàtics (Verheul, 2006, pàg. 52).

2.3. Altres actuacions

A part de les estratègies descrites en els apartats anteriors, també hi ha altres pràctiques que solen incloure's juntament amb les estratègies, encara que algunes no s'apliquin estratègicament sinó per defecte, per falta d'altres opcions assequibles:

Els museus informàtics: aquesta opció preveu el manteniment de "museus" de programari i maquinari antics –tots els models, més còpies i peces de reparació– per a garantir la lectura de dades en el seu entorn original tal com van ser concebuts. De moment, aquesta opció no té resposta per als interrogants següents:

- Qui guardarà i mantindrà aquests equips?
- Com se'n garantirà el correcte funcionament?
- Com s'integraran al sistema vell els nous suports d'emmagatzemament o dispositius de lectura o reproducció? És ineludible la seva incorporació a causa de la necessitat de fer còpies periòdiques per a refrescar les dades i de la necessitat dels usuaris d'utilitzar els recursos en un futur.

La impressió en suports analògics: els promotors d'aquesta opció, una mica "retrògrada", defensen la transferència d'informació guardada en suports digitals a altres de tradicionals com el paper no àcid o el microfilm, ja que aquests suports tradicionals són més estables. A més, ja hi ha en moltes institucions els coneixements i els procediments necessaris per a garantir la preservació d'aquest tipus de suport. En una enquesta feta al Regne Unit per la Digital Preservation Coalition el novembre del 2005, es va demostrar que el 50% de les organitzacions que hi participaven continuaven imprimint sobre paper els seus documents digitals per a assegurar-ne la conservació (Waller i Sharpe, 2006, pàg. 15). Però, malgrat ser una tècnica encara adequada per a documents estàtics com text i imatges fixos, la presència en recursos digitals d'elements multimèdia, enllaços hipertext i dades dinàmiques limita la utilitat d'aquesta opció.

La preservació en línia: alguns experts proposen la preservació en línia per evitar l'emmagatzemament d'informació en suports físics amb una vida funcional molt limitada. Argumenten que:

- el cost de dispositius d'emmagatzematge en línia continua descendint,
- els formats com l'HTML i l'XML duraran més temps que els formats de programes de propietat, i
- s'actualitzen sobre la marxa de manera natural.

A favor de la microfilmació

Un document preparat el 2002 a càrrec del Fòrum DLM celebrat a Barcelona va postular que l'única manera de garantir la preservació dels materials digitals estàtics a llarg termini –concepte definit en aquest informe com a més enllà de deu anys– és mitjançant la seva transferència a microformes. Aquestes reproduccions gràfiques podran servir de prova de l'autenticitat del recurs digital, ja que, a diferència de l'original digital, no es podran manipular. Sembla evident que aquest procés no és aplicable a recursos dinàmics, interactius o amb elements de multimèdia i dista molt de la promesa de la facilitat d'ús i els arxius digitals (*Availability & preservation...*).

Sembla l'opció per defecte de molts centres. De tota manera, no elimina la necessitat de fer migracions en algun moment a futurs formats encara desconeguts. Tampoc no evita el risc de destrucció física d'un arxiu emmagatzemat en un únic servidor.

L'arqueologia digital: de fet, aquesta opció, més que una "estratègia", s'assembla a una absència d'estratègies. Es tracta de deixar la recerca de solucions a les mans d'especialistes, els "arqueòlegs digitals" de futures generacions, ja que actualment els nostres coneixements no permeten garantir la supervivència de tot el que es produeix avui en dia.

També es parla de "l'arqueologia digital" en parlar de rescats de recursos no recuperables per altres mitjans. L'exemple esmentat molt sovint és la recuperació de dades del Domesday Project gravat per la British Broadcasting Company (BBC) en discos làser al començament dels anys vuitanta. Aquest esforç va durar més d'un any i va ocupar els esforços d'un equip d'especialistes.

El "no fer res": donada la varietat dels riscos que enfronten els recursos digitals i la impossibilitat avui en dia d'assegurar-ne la perdurabilitat, no és estrany que algunes institucions escullin no fer res, esperant unes solucions definitives. Encara que es pugui entendre aquesta actitud, també és cert que "no fer res" perjudica encara més els documents digitals, ja que, com afirmen els experts, una intervenció primerenca és un pas important cap a la preservació.

En resum, no es tracta d'escollir un únic mètode per a una institució, sinó d'emprar una combinació d'aquestes opcions en funció del contingut i el format del recurs, de la capacitat institucional i de les necessitats previstes per a accedir al recurs en el futur.

De quines novetats disposarem en el futur?

No podem imaginar les novetats que ens aportarà la tecnologia en el futur. Per exemple, fa cinc anys qui hauria pensat que un telèfon pogués treure fotos o que es podria visualitzar casa nostra des d'un satèl·lit mitjançant un programa de distribució gratis com GoogleEarth? Les tecnologies del futur possiblement podran desxifrar els documents digitals d'una manera que encara supera les nostres possibilitats actuals.

3. Els dipòsits digitals i la gestió del material custodiat

La qüestió sobre on i com es poden emmagatzemar els recursos digitals és senzilla i complexa alhora:

- senzilla, perquè si es vol preservar la informació digital, òbviament cal emmagatzemar-la (i així ho fan, d'una manera o altra, la majoria de les institucions);
- complexa, perquè encara queden incògnites relacionades amb la manera d'assegurar la permanència de la informació emmagatzemada perquè futures generacions d'usuaris puguin localitzar-la, llegir-la i utilitzar-la. Afegit a aquesta incertesa, els recursos econòmics, humans i materials que previsiblement seran necessaris per a garantir la preservació actualment són fora de l'abast de moltes institucions.

El progrés viscut en l'entorn de la preservació digital durant els cinc últims anys ha conduït a un cert optimisme quant a la resolució –o almenys la contenció– dels problemes tècnics, sense infravalorar els reptes que encara queden per superar. En l'enquesta sobre la preservació del patrimoni digital duta a terme per Geser i Pereira, dos experts entrevistats afirmen que la tecnologia ja no és ni l'arrel del problema, ni la font de la solució:

"No fa falta parlar de la tecnologia com a solució, ni de bon tros com la qüestió principal. Fins i tot no sabem què fer amb la tecnologia ja disponible i, a més, els informàtics no estan capacitats per a resoldre els problemes." (pàg. 58-59)

"Els grans salts per a avançar la gestió de la preservació digital a llarg termini no estan bloquejats per qüestions tecnològiques: «de fet, tecnològicament, podríem desenvolupar les solucions ara mateix»." (pàg. 60)

(Geser i Pereira, 2004)

Els problemes que actualment semblen més difícilment solubles estan relacionats amb la capacitat política, administrativa i econòmica de les institucions per a sostenir la infraestructura i mantenir les activitats necessàries per a assegurar que els recursos continuïn accessibles per a futures generacions. Si de vegades s'identifica la falta de recursos econòmics com un obstacle molt important, també és cert que es pot aconseguir molt sense grans inversions. En canvi, no es pot garantir la preservació dels recursos digitals a llarg termini sense la implicació de la institució –també a llarg termini– per a l'elaboració i aplicació de polítiques que permetin una actualització contínua de la infraestructura i l'adaptació dels processos, d'acord amb l'evolució dels entorns intern i extern: els canvis en la tecnologia, el marc legal, els pressupostos disponibles, etc.

El suport institucional sol involucrar també canvis de la cultural institucional la introducció dels quals pot resultar difícil i complexa: canvis de personal, de responsabilitats, de procediments i, sens dubte, d'actituds. Per aquesta raó la implantació de polítiques i procediments per a facilitar i promoure la preservació digital en les institucions avança de manera desigual, segons la capacitat, els recursos i la voluntat de cada una d'elles.

És per això que moltes institucions prefereixen esperar que hi hagi més experiència col·lectiva i que s'elaborin normes abans d'assumir formalment la responsabilitat de la preservació digital.

L'any 1996, l'informe pioner sobre la preservació digital i les biblioteques, *Preserving digital information*, va recomanar la creació de centres especialitzats per a la preservació de recursos digitals a llarg termini. Però no va ser possible iniciar aquests centres en aquell moment, a causa de la falta de coneixements, experiències i criteris que servissin d'orientació i la falta d'eines per a dur a terme la tasca. No obstant això, la idea de l'establiment de centres especialitzats i autoritzats per a fer aquestes activitats continuava atraient adeptes. Com a fruit dels esforços dels cinc últims anys hi ha hagut importants avenços, sobretot en la definició, el disseny i la implementació dels dipòsits dedicats a la preservació a llarg termini.

3.1. Dipòsits digitals fiables

Una fita en aquest camí va ser la publicació el 2002 de *Trusted digital repositories: attributes and responsibilities*, elaborat per un grup de treball posat en marxa per dos grans consorcis de biblioteques, Research Libraries Group (RLG) i OCLC², i basat en el protocol OAIS, descrit en la secció següent. L'informe va proposar les característiques dels dipòsits que assumiran les responsabilitats – també identificades a l'informe – de mantenir i gestionar les col·leccions digitals, sobretot les d'institucions d'investigació, tant grans com petites. Segons l'informe, la missió d'aquests dipòsits "fiables" és garantir que els recursos digitals gestionats per ells es mantinguin accessibles a llarg termini, de manera fiable, per a una determinada "comunitat designada d'usuaris" (*designated user community*).

Aquesta definició inclou dos conceptes primordials que recorren tota la documentació sobre els dipòsits digitals: fiabilitat i comunitat designada d'usuaris.

Fiabilitat: des de fa segles la societat en general ha pogut beneficiar-se de l'existència de mostres del patrimoni cultural i d'altres recursos valuosos d'èpoques passades, ja que certes institucions de "memòria cultural" –les biblioteques, els arxius i els museus– han estat capaços de dur a terme aquesta activitat. No se sol dubtar de l'autenticitat dels documents guardats en una biblioteca o dels quadres exposats a les parets d'un museu (encara que de tant en tant puguin sorgir casos dubtosos). Segurament un analitzaria amb més

⁽²⁾El juny del 2006, aquests dos consorcis van anunciar la seva fusió.
<http://www.oclc.org/news/releases/200625.htm>

perspicàcia les obres a les mans de galeries o de col·leccionistes particulars. La raó és, senzillament, que les institucions de memòria cultural han demostrat la seva fiabilitat al llarg de dècades i, en alguns casos, de segles.

Gràcies a la seva tasca, avui en dia encara es pot gaudir de les obres clàssiques del passat; també s'espera que, gràcies a elles o a altres institucions que en prenguin el relleu, les creacions intel·lectuals de la nostra època arribin a les mans dels nostres descendents. Perquè això sigui així amb les creacions digitals, caldrà trobar la manera de demostrar o provar la fiabilitat del següent:

- la institució responsable de preservar el recurs custodiat,
- els mètodes emprats per a dur a terme les activitats necessàries per a la preservació i
- els recursos custodiats en el moment de ser recuperats.

Expressat des de la perspectiva de qui haurà de fiar-se de qui (persones o institucions), l'informe de l'RLG i l'OCLC identifica els tres agents següents:

- els que hi estan més directament implicats han de poder confiar en la institució que assumeixi la responsabilitat de la preservació;
- la institució responsable de la preservació ha de poder confiar en les entitats a les quals possiblement traspassi una part de la responsabilitat: per exemple, la gestió tècnica contractada a una empresa de serveis o un consorci de biblioteques;
- la comunitat d'usuaris, avui i en el futur, ha de poder confiar en les dades preservades i recuperades.

Segons els autors de l'informe, un punt crític és que el dipòsit i, sens dubte, la institució a què pertany puguin demostrar que són dignes de confiança (*trustworthiness*) i capaços de demostrar la seva fiabilitat (*reliability*) al llarg del temps, sigui quina sigui l'estratègia o el mètode de preservació emprat, ja que aniran canviant segons l'evolució del mercat.

Però abans de poder comprovar la fiabilitat institucional, cal identificar i definir els conceptes i criteris bàsics que serviran per a demostrar-la. És per això que l'informe sobre els dipòsits fiables proposa un marc per a la definició d'allò que s'exigeix a un dipòsit fiable i també per a la normalització dels conceptes i la terminologia. Disposar d'una terminologia normalitzada pot semblar un pas modest, però representa una peça clau per a la interoperabilitat entre els dipòsits en el futur.

Els conceptes i criteris definits hauran de ser avalats per tots els *stakeholders*; és a dir, aquells agents relacionats directament o indirectament amb la preservació o amb el recurs preservat: creadors, distribuïdors, gestors, subvencionadors o usuaris (aquests són l'enfocament de l'apartat següent).

Un altre concepte bàsic té a veure amb els usuaris actuals i futurs; és a dir, la *designated user community*. Si l'objectiu de la preservació dels recursos és que siguin accessibles i utilitzables en el futur, això pressuposa l'existència d'un grup d'usuaris específic, aquells que sabran interpretar i utilitzar la informació preservada. Qui són i com són identificats o seleccionats?

Una premissa bàsica de la preservació tradicional és saber per a qui es preserva un document. Quan es tracta de recursos digitals, és encara més important, ja que pot condicionar aspectes clau del procés. Qui són els usuaris d'avui i qui seran els del futur? Alguns grups seran molt amplis, heterogenis, i, fins i tot, indefinits, com seria el cas dels ciutadans d'un país sencer que formen la comunitat servida per la biblioteca nacional. D'altres seran grups reduïts, consolidats, amb membres que s'autoidentifiquin o s'autolimitin; seria el cas dels investigadors que treballen en un camp concret, per exemple els astrofísics. Els responsables de cada dipòsit han d'identificar la seva pròpia comunitat d'usuaris i preveure com utilitzaran els recursos preservats; aquesta informació serà necessària per a la presa de decisions, per exemple sobre les propietats essencials que cal preservar o les estratègies tècniques que cal emprar per a garantir la preservació del recurs.

A més de mantenir la infraestructura tècnica adequada per a garantir l'accés futur a la informació digitalitzada, es preveu que els dipòsits fiables també hauran de complir determinades obligacions, com ara:

- garantir la integritat i l'autenticitat del recurs,
- vigilar la protecció dels drets intel·lectuals,
- limitar l'accés a usuaris autoritzats,
- mantenir les eines necessàries per a identificar i recuperar la informació, i
- assegurar la presència de metadades normalitzades, que sempre acompanyaran la informació preservada.

L'informe original d'RLG i OCLC es va concebre com un primer pas i, com a tal, va recomanar una sèrie d'accions que facilitessin la creació de dipòsits fiables i d'un sistema per a la seva certificació. Les accions clau recomanades el 2002 van ser:

- desenvolupar un procés per a la certificació de dipòsits digitals;
- investigar i crear instruments per a identificar les propietats essencials dels recursos digitals que s'han de preservar;
- investigar i desenvolupar models de xarxes i de serveis cooperatius de dipòsits;

- desenvolupar sistemes per a la identificació única i persistent dels recursos (*unique, persistent identification*); aquest fet també es considera un element clau per a la preservació a llarg termini;
- investigar i difondre informació sobre la complexa relació entre la preservació digital i els drets de propietat intel·lectual;
- identificar les estratègies tècniques que ofereixin la millor garantia de l'accés continuat;
- definir les metadades elementals necessàries per a la gestió a llarg termini i desenvolupar instruments per a generar i/o extreure automàticament la màxima quantitat possible.

Els dos consorcis que van encarregar l'estudi original –RLG i OLC– continuen plenament lliurats a la recerca de solucions per a la preservació digital. Com a exemple, cada un s'ha dedicat a dur a terme algunes de les recomanacions citades en el paràgraf anterior: per exemple, la certificació de dipòsits, per part d'RLG, i la identificació i normalització de metadades, per part d'OCLC. Aquestes iniciatives, entre d'altres, estan descrites més endavant.

D'altra banda, diverses organitzacions –sobretot grups de biblioteques treballant col·lectivament– s'han involucrat en iniciatives que cerquen solucions a altres aspectes problemàtics que van ser identificats en l'estudi original. Les iniciatives, algunes de comentades en altres seccions, cobreixen temes com ara:

- propietats essencials,
- dipòsits cooperatius,
- identificadors persistents,
- estratègies tècniques i
- propietat intel·lectual.

3.2. *Open Archival Information System*

La proposta d'RLG i OCLC per a la creació de dipòsits fiables, plasmada en l'informe *Trusted digital repositories*, es basa en el model referencial (*reference model*) Open Archival Information System (OAIS). De fet, la seva proposta representa l'aplicació de l'OAIS en un sector determinat –el de les biblioteques– tal com estava previst quan es va elaborar el model original.

El model OAIS va ser desenvolupat per un comitè consultiu de la nord-americana National Aeronautics and Space Administration (NASA) amb la col·laboració d'altres agències nacionals aeroespacials i es va convertir posteriorment en la norma internacional ISO 14721:2003: Open Archival Information System (OAIS). A més del cas de l'adaptació feta per les biblioteques, el model OAIS també ha estat adoptat en altres àmbits a escala internacional.

Què és un model referencial?

La definició oferta pel mateix protocol OAIS és la següent: "Un marc per a entendre les relacions importants entre les entitats d'un entorn i per al desenvolupament d'estàndards o especificacions coherents per a donar suport a aquest entorn. Un model de referència es basa en un nombre petit de conceptes unificadors i es pot fer servir com a base per a l'educació i per explicar estàndards a algú no especialista." (*Reference model.*, pàg. 1-12)

Agències aeroespacials i preservació digital

La preocupació per la preservació digital mostrada per agències aeroespacials nacionals – alienes al món de les biblioteques– neix de les quantitats ingents d'imatges i altres tipus d'informació digital generades pels satèl·lits i gestionades per elles per a assegurar-ne la recuperació i usabilitat en el futur. Segons la pàgina web de la NASA, el telescopi espacial Hubble diàriament arxiva entre 3 i 5 GB de dades i difon entre 10 i 15 GB a astrònoms de tot el món. Aquesta web ofereix accés lliure a algunes de les imatges més espectaculars captades i que aporten vistes de la Terra i d'altres racons de l'univers.

<http://www.nasa.gov/multimedia/imagegallery/index.html>

Si el motiu per a l'elaboració d'aquest model va ser la definició –en línies generals– dels requeriments per a la preservació de recursos digitals a llarg termini, el primer pas que es va fer va ser normalitzar la terminologia relacionada amb l'entorn, els components funcionals i els "objectes d'informació" (*information objects*) que calia preservar. Tal com ja es va afirmar en la secció prèvia, la definició d'elements i conceptes bàsics afavoreix la interoperabilitat en el futur entre centres dedicats a la preservació digital. En paraules del consultor britànic H. M. Gladney, OAIS permet que "els usuaris sàpiguen que «són a la mateixa pàgina»: és a dir, que parlen de la mateixa cosa..." (Gladney, 2006).

El model OAIS no ofereix un disseny específic d'arquitectura, sinó que serveix com a "full de ruta" que facilita la comunicació entre institucions, mitjançant la normalització terminològica i conceptual, per a reduir la possibilitat de malentesos. Gladney segueix la seva descripció així: la norma

"conté pistes implícites sobre una possible arquitectura, però no són suficients per a jutjar si les implantacions són les adequades."

Per a il·lustrar aquest nivell de generalització, aquest autor suggereix l'exemple següent de la definició de *cotxe*:

"Un vehicle de quatre rodes, capaç de transportar entre dues i vuit persones amb una possible velocitat d'uns 130 km/h... amb espai per a portar una quantitat modesta d'equipatge i amb prestacions perquè el viatge sigui més agradable (ràdio, aire condicionat, seients còmodes), pneumàtics..."

Per als que no hagin vist mai un cotxe, aquesta descripció ofereix una idea general i també introdueix un vocabulari que es refereix als elements i conceptes bàsics que, si fos adoptat per tots, serviria com a base d'una terminologia normalitzada. De manera semblant, la norma OAIS ofereix un esbós d'un dipòsit digital i introdueix la terminologia bàsica per a referir-s'hi. Però tant en el cas del cotxe com en el de l'OAIS, les definicions disten molt d'oferir un disseny concret per a desenvolupar un producte operatiu, sigui un vehicle o un dipòsit.

3.2.1. Components i funcionalitat de l'OAIS

La definició de cotxe suggerida per Gladney en descriu en termes generals l'estructura i funcionalitat bàsiques. De manera semblant, les "pistes implícites" de l'OAIS esbossen una arquitectura que inclou sis àrees funcionals, de les quals quatre són activitats a què se sotmet el material dipositat:

- la seva incorporació (*ingest*) al dipòsit,

- el seu emmagatzemament a llarg termini (*archival storage*),
- la seva gestió (*data management*) i
- el seu accés/difusió (*access/dissemination*).

Les dues restants tenen a veure amb el funcionament del dipòsit:

- planificació per a la preservació (*preservation planning*) i
- gestió del dipòsit (*archive administration*).

Es refereix al recurs que es pretén preservar com a "paquet d'informació" (*information package- IP*) que inclou tant l'objecte digital com les metadades –imprescindibles– que hi estan associades. A l'IP se li assignen qualificatius per a referir-se a les diferents manifestacions del recurs, segons l'àrea funcional en la qual es trobi en un moment determinat:

- *Submission information package* (SIP): l'objecte, juntament amb les metadades associades, tal com arriba al dipòsit.
- *Archival information package* (AIP): el recurs dipositat com a SIP, amb les modificacions necessàries, s'encapsula dins de l'AIP i s'emmagatzema en forma de cadena de bytes (*bytestream*) per a la seva gestió i preservació. Les metadades són imprescindibles perquè la cadena de bytes pugui convertir-se, en qualsevol punt del futur, en un recurs digital comprensible i utilitzable.
- *Dissemination information package* (DIP): el recurs tal com es lliura a l'usuari, que tindrà una versió de l'objecte original possiblement canviat pels processos de gestió, migració, etc. El DIP conté, a més de l'objecte en si mateix, les metadades i qualsevol altra informació o aplicació necessària perquè l'usuari sàpiga utilitzar-lo.

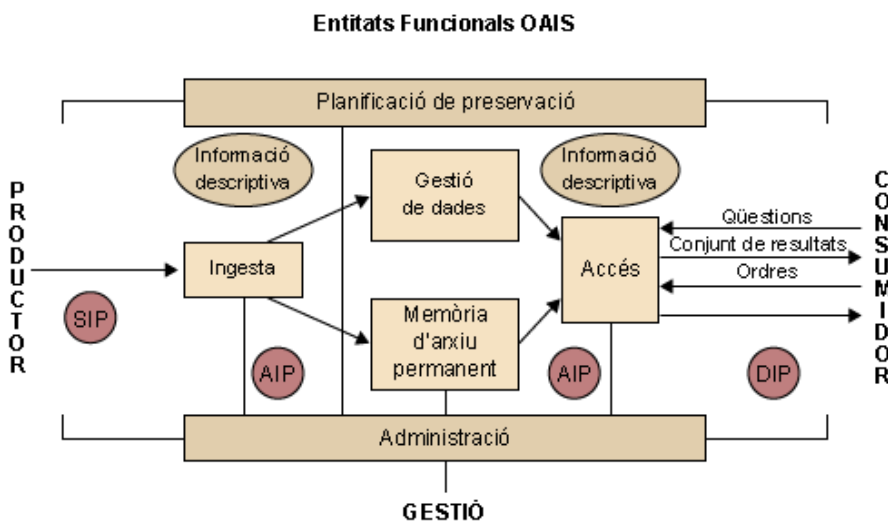
L'OAIS reconeix dos tipus de metadades necessàries per a la preservació del recurs custodiat:

- *representation information* (RI): aquesta informació de representació conté dades tècniques necessàries per a convertir l'objecte digital a altres formats segons la necessitat, per a permetre que sigui recuperable en el futur;
- *preservation description information* (PDI): aquesta informació conté dades descriptives –com les típicament incloses en una descripció bibliogràfica– que permeten la identificació i recuperació de l'objecte en el futur; també se n'hi incorporen d'altres que cobreixen aspectes com els drets d'autor, restriccions de l'accés, etc. Les dades incloses en la PDI s'agrupen en les categories³ següents, resumides de manera molt general:
 - referència (*reference*): informació que identifica com a objecte únic el recurs custodiat;
 - procedència (*provenance*): informació sobre la història del recurs;

⁽³⁾Els articles de Candas i de Salvador i Ruiz, esmentats en la bibliografia, ofereixen més detalls sobre el significat de cada una d'aquestes categories.

- fixesa (*fixity*): informació per a comprovar l'autenticitat i integritat del recurs;
- context (*context*): informació que descriu la relació del recurs amb d'altres.

Aquesta combinació de processos i productes s'il·lustra en el gràfic següent (*Reference model...*, pàg. 4-1):



SIP = paquet informatiu de presentació

AIP = paquet informatiu d'arxiu

DIP = paquet informatiu de divulgació

<http://public.ccsds.org/publications/archive/650x0b1.pdf>

L'informe *Trusted Digital Repositories* identifica algunes de les operacions habituals relacionades amb les àrees funcionals de l'esquema OAIS: incorporació, emmagatzemament, gestió, planificació per a la preservació i gestió del dipòsit i accés/difusió.

Incorporació:

- Assignar o validar un identificador únic.
- Seleccionar l'estratègia tecnològica basada en la definició de les propietats essencials del recurs.
- Convertir el recurs i les seves metadades en una cadena de bits.
- Elaborar la RI necessària.
- Verificar o elaborar la PDI.

Emmagatzemament:

- Passar els recursos de l'àrea d'incorporació a la d'emmagatzemament permanent.
- Gestionar el sistema d'emmagatzematge en què s'han d'incloure les activitats relacionades amb la seguretat del sistema (còpies de seguretat, etc.).

La paraula *archive*

És interessant tenir en compte l'eliminació de la paraula *archive* que figura en la norma OAIS, que va ser reemplaçada per *repository* en l'informe de l'RLG i l'OCLC, ja que els autors consideraven que *archive* era massa ambigua i, en l'entorn de les institucions culturals, es podia confondre amb el concepte tradicional arxivístic.

- Refrescar els suports d'emmagatzemament: inclou les comprovacions sobre l'estat dels suports i de la integritat de les dades.
- Subministrar tota la informació necessària perquè es pugui accedir al recurs i difondre'l.

Gestió de les dades:

- Controlar-ne l'accés.
- Fer el seguiment de les sol·licituds dels usuaris.
- Mantenir la informació següent:
 - els perfils dels usuaris;
 - les dades relacionades amb la seguretat dels recursos: noms d'usuari, contrasenyes, certificació digital, etc.;
 - les estadístiques sobre el funcionament del sistema, i
 - les dades de comptabilitat.

Planificació per a la preservació:

- Fer un seguiment de:
 - la comunitat designada d'usuaris,
 - les novetats tecnològiques i
 - les propietats essencials dels recursos custodiats.
- Desenvolupar:
 - estratègies i pautes per a assegurar l'accés continuat i
 - plans per a la transferència i migració de dades.

Gestió del dipòsit:

- Negociar els acords amb els creadors/proveïdors dels recursos.
- Revisar els procediments vigents.
- Desenvolupar i mantenir actualitzades les polítiques i pautes del dipòsit.
- Oferir suport als usuaris.
- Mantenir contactes amb:
 - l'administració institucional i
 - altres dipòsits.

Accés/difusió:

- Comprovar la integritat del recurs lliurat a l'usuari.
- Verificar el nivell d'autorització de l'usuari per a accedir al recurs sol·licitat.

3.2.2. Identificació automàtica dels elements tècnics

Quan s'incorporen recursos nous en el dipòsit, una de les primeres accions és la comprovació del format, mitjançant indagacions com les següents:

- Quin és el format –i versió del format– del recurs?
- Segueix l'especificació del format al 100%?
- Conté altres formats encapsulats dins del format principal?
- Quines de les propietats del format es troben manifestes en aquest recurs?

En el cas del dipòsit directe per part d'una persona o una entitat, se suposa que l'encarregat de dipositar el recurs subministrà aquesta informació. Però poden sorgir obstacles: de vegades el mateix autor desconeix els detalls tècnics o hi apareixen dades errònies. Per exemple, el projecte AIHT, que analitza la problemàtica relacionada amb la incorporació, exportació i transferència de recursos digitals entre diferents institucions, ha descobert casos d'extensions errònies dels fitxers importats: per exemple, s'han detectat fitxers amb l'extensió .jpg que en realitat contenien imatges en format GIF (Shirky, 2005).

En el cas de la incorporació massiva de recursos mitjançant un procés de recollida o *harvesting* –com en el cas dels arxius de captació de material web–, no hi ha persones o entitats dipositants que puguin facilitar aquesta informació i, com quedava exposat en el paràgraf anterior, algunes indicacions formals poden ser equívokes.

La informació sobre el format s'incorpora en les metadades anomenades *informació de representació* (RI), tal com estableix el model OAIS. Aquestes dades són imprescindibles per a les operacions de preservació, ja que, a més d'identificar els aspectes tècnics del recurs necessaris per a assegurar-ne la futura usabilitat, també poden condicionar la selecció d'estratègies de preservació i, fins i tot, la decisió sobre si podrà assumir-se, o no, la responsabilitat de la seva custòdia.

Aquesta informació de representació, necessària per a interpretar correctament el format d'un recurs digital, conté dos nivells:

- informació genèrica sobre el format en si mateix i
- informació més específica sobre l'aplicació del format en el recurs dipositat.

Un exemple de les dades a dos nivells corresponent a un fitxer TIFF és l'ofert per Stephen Abrams de la Universitat de Harvard i coordinador del projecte JHOVE (Abrams, 2004):

- El primer nivell fa referència a les propietats del format TIFF en general: la gamma de colors possibles que pot tractar el format o els algoritmes de compressió que pot reconèixer.

- El segon nivell fa referència al recurs dipositat: especifica exactament quines de les propietats del TIFF es manifesten en aquest recurs, la paleta de colors usada i la compressió emprada.

És evident que les dades del primer nivell són sempre iguals per a tots els centres i per això és important l'existència del Global Digital Format Registry (GDFR), descrit en l'apartat 1 ("La informació digital: els seus elements i les seves amenaces"), la finalitat del qual és recopilar i subministrar informació completa i fiable sobre tots els formats representats en els dipòsits digitals.

El segon nivell d'informació és específic per al recurs dipositat i és generat pel mateix dipòsit, basant-se en les dades disponibles quan el recurs es diposita. Aquestes poden ser incompletes o equívokes, la qual cosa generaria metadades invàlides, tret que es revisessin individualment –un procés laboriós i, per aquest motiu, costós.

És per això que el projecte JHOVE va proposar la creació d'un sistema que automatitzés els processos relacionats amb la comprovació del format en el qual hi ha codificats els recursos en el moment de la seva incorporació a un dipòsit digital. D'aquesta manera se simplifica el procés d'integració, que així és menys costós per al centre i, alhora, es garanteix la generació de metadades vàlides, la qual cosa contribueix a una fiabilitat més gran, tant de la gestió com dels recursos recuperats.

Els passos duts a terme per l'aplicació JHOVE corresponen a la identificació, la validació i la descripció de les característiques del format present en el recurs dipositat:

- identificar el format del recurs, incloent-hi la seva versió o subversió;
- validar que el contingut realment segueixi l'especificació del format declarat;
- determinar el nivell de conformitat amb els requeriments formals, i
- caracteritzar els elements propis del format que es manifesten en el recurs dipositat.

3.3. Avaluació dels dipòsits fiables

Tal com s'ha comentat al principi d'aquesta secció, l'estudi original de l'OCLC i l'RLG va proposar els atributs i responsabilitats dels dipòsits digitals fiables, i una de les seves recomanacions va ser la de desenvolupar un procediment que permetés la certificació d'aquests dipòsits. Aquesta certificació es considera imprescindible perquè els dipòsits puguin demostrar la seva fiabilitat a altres implicats en el cicle vital dels recursos:

- als creadors i/o titulars dels recursos dipositats,
- a la comunitat designada d'usuaris que eventualment recuperaran els recursos,

- als que financen les operacions i
- a altres entitats que col·laboren d'una manera o una altra en les activitats de preservació.

Però possiblement els més interessats seran els mateixos dipòsits que necessitaran comprovar si les seves pròpies operacions satisfan els criteris establerts i, en cas negatiu, rectificar les faltes detectades. L'adequació de les activitats no solament es jutja per la seva capacitat tècnica, sinó per la capacitat i voluntat institucional de dur a terme les operacions necessàries actualment i de sostenir-les –econòmicament– a llarg termini.

Malgrat la importància reconeguda d'un sistema de certificació, no és possible crear-lo sense que existeixin prèviament:

- uns criteris avaluable per a saber si el dipòsit compleix els atributs i les responsabilitats establerts i, a més,
- un mecanisme per a mesurar el nivell de compliment.

Per a remeiar la falta d'aquests instruments, s'han dedicat molts esforços en aquests últims anys a determinar la millor manera d'establir un sistema d'avaluació, mesurament i certificació dels dipòsits fiables.

3.3.1. *Audit checklist for the certification of trusted digital repositories*

Com a primer pas, el 2005 l'IRLG, coeditor de l'estudi original sobre els dipòsits digitals fiables, juntament amb l'arxiu nacional nord-americà, NARA, van elaborar una llista provisional de criteris que podrien utilitzar-se per a avaluar i certificar l'actuació dels dipòsits: *Audit checklist for the certification of trusted digital repositories* (encara en versió provisional a les acaballes del 2006). El seu propòsit va ser elaborar un mecanisme que facilités la identificació dels centres "capaços d'emmagatzemar, migrar i donar accés a fons digitals, de manera fiable". La introducció d'aquest document insisteix que el procés de certificació també ha de disposar, més enllà dels criteris, d'eines per a l'avaluació –tant interna com externa– de les operacions.

Els criteris proposats en aquest document permeten avaluar els dipòsits atenent quatre àrees operacionals:

- organització;
- funcions, processos i procediments;
- comunitat d'usuaris i usos previstos, i
- tecnologia i infraestructura tècnica.

Dins de cada d'aquestes àrees, hi ha subcategories que se centren en aspectes més específics. Per exemple, en el cas de l'àrea operacional, "funcions, processos i procediments", es proposen criteris per a avaluar el nivell d'ús de determinades funcions (basades en OAIS), com ara:

- la incorporació del recurs;
- la gestió del recurs emmagatzemat;
- la planificació i les estratègies de la preservació;
- la gestió de les dades (principalment la seva descripció), i
- la gestió de l'accés.

Per a donar una millor idea d'aquests criteris, se'n resumeixen a continuació alguns de relacionats amb la funció *c* (planificació i estratègies de la preservació), dels quals el dipòsit haurà de demostrar el compliment:

- El dipòsit té documentades les seves estratègies de preservació.
- El dipòsit realitza les estratègies per a l'emmagatzemament i la migració del recurs (AIP).
- El dipòsit utilitza i aplica la informació tècnica (RI) disponible en els registres internacionals.
- El dipòsit comprova contínuament la integritat del recurs custodiat.
- El dipòsit té un mecanisme per a detectar quan un format comença a ser inutilitzat i s'aproxima l'amenaça de quedar obsolescent.
- El dipòsit té un mecanisme per a modificar la seva política de preservació com a resultat del seguiment de les tendències externes o els canvis interns.
- El dipòsit pot demostrar l'èxit de la planificació realitzada.

Una prova de l'impacte potencial d'aquesta eina d'avaluació és l'interès que ha despertat en organitzacions prestigioses a escala internacional. Actualment l'esquema està en una fase de revisió i validació, coordinada pel Center for Research Libraries (CRL), per a conèixer l'adequació de la seva aplicació en dipòsits operatius. Els participants en aquesta revisió avaluen els criteris proposats en l'*Audit checklist* mitjançant una autoavaluació dels seus propis procediments, serveis i recursos institucionals. Les institucions avaluadores són les següents:

- InterUniversity Consortium for Political and Social Research
- Koninklijke Bibliotheek
- LOCKSS
- Portico

És evident que l'elaboració i validació d'aquests criteris permetran que el col·lectiu bibliotecari avanci cap a un sistema de certificació. Però encara falten altres passos intermedis perquè es pugui dur a terme: per exemple, no hi

ha pautes que indiquin a la biblioteca com ha de fer els processos avaluable. Seguint l'exemple anterior dels criteris relacionats amb la planificació i les estratègies de la preservació, els dipòsits hauran de saber:

- Quina és la manera recomanada de comprovar la integritat del recurs custodiat?
- Amb quina freqüència cal fer aquesta comprovació?

L'*Audit checklist* tampoc no suggereix la manera com el dipòsit pot mostrar que compleix el criteri, per exemple:

- Quins indicadors s'exigeixen per a demostrar la integritat dels recursos?
- Quines proves es requereixen per a demostrar el nivell d'èxit dels esforços de planificació?

Digital Curation Centre: en un intent de remeiar la falta d'especificitat de l'*Audit checklist*, el Digital Curation Centre (DCC) ha proposat tres tipus d'evidència que servirien per a demostrar el nivell de compliment dels criteris per part dels dipòsits digitals, aspirants a la certificació de fiabilitat: es tracta de l'evidència en forma de documents, testimonis i observació. Quant a l'evidència documental, per exemple, el DCC suggereix quins documents podrien oferir-se per a demostrar l'adequació de les actuacions i quina informació haurien d'incloure. El dipòsit hauria d'elaborar i mantenir actualitzada la documentació com a part íntegra de les seves operacions; alhora, serviria com a prova de la bona gestió dels recursos que es podrien presentar durant un procés d'avaluació. Entre els documents clau s'identifiquen els següents:

- la missió del dipòsit,
- els acords amb els qui dipositen els recursos (autors, editors, etc.),
- les descripcions de càrrecs laborals i de responsabilitats corresponents,
- l'organigrama institucional,
- els informes financers anuals,
- l'anàlisi de riscos,
- el manual de procediments,
- els fluxos de treball, i
- l'arquitectura tècnica.

El DCC està motivat en aquesta tasca per la seva pretensió d'oferir un futur servei de certificació per als dipòsits digitals del Regne Unit. Una vegada posat en marxa, se suposa que augmentarà el nombre de dipòsits disposats a certificar-ne la fiabilitat (Ross i McHugh, 2006).

ECHO DEPOSITORY: un altre projecte també pretén aplicar l'*Audit checklist*, però amb un propòsit diferent. Es tracta de l'ECHO DEPOSITORY, dut a terme per un equip format per la Universitat d'Illinois i l'OCLC, entre d'altres, que forma part del Programa nacional de preservació nord-americana, NDIIPP, coordinat i subvencionat per l'LC. La finalitat del projecte és el desenvolupament

d'una metodologia i el disseny d'unes pràctiques per a la captura de material web. Una de les activitats que s'han dut a terme ha estat l'elaboració d'un marc per a avaluar el programari de gestió de materials digitals com, per exemple, DSpace, Eprints, Fedora o Greenstone. Es va elegir l'*Audit checklist* com a instrument d'avaluació i després es va adaptar perquè s'adeqüés més al canvi d'enfocament: l'avaluació de programari en comptes de les operacions dels dipòsits (Kaczmarek i altres, 2006).

L'equip encarregat d'aquesta adaptació ofereix l'exemple següent de les modificacions efectuades; en aquest cas es tracta d'una sèrie d'anotacions per a precisar millor les funcions necessàries per a satisfer un dels criteris de l'*Audit checklist*.

Criteri original: el dipòsit facilita la difusió de còpies autèntiques de l'original o d'objectes que estableixen la seva relació amb l'original (B5.6).

Les anotacions a aquest criteri introduïdes per ECHO DEPOSITORY per a avaluar el programari de gestió de material digital són les següents:

- Serà fàcil per a l'usuari validar l'objecte que se li hagi lliurat (DIP) mitjançant les sumes de verificació (*checksums*)? Té accés a les sumes de verificació? Són segures les sumes de verificació?
- El programari admet les metadades de procedència? S'inclouen aquestes dades en la versió lliurada a l'usuari (DIP)? Quines dades sobre procedència hi estan incorporades?
- El programari admet el control de múltiples versions? Es modifica la informació referencial (RI) de les versions originals quan es modifiquin o quan es dipositi una versió nova?

Sens dubte, quan estigui disponible aquest instrument serà molt útil, ja que qualsevol programari adquirit per a la gestió a llarg termini de recursos digitals haurà de demostrar la seva capacitat de dur a terme les funcions que s'exigeixen al dipòsit. Però, tal com adverteixen els membres de l'equip d'ECHO DEPOSITORY, si encara es discrepa sobre el concepte de "dipòsit digital", queda molt lluny un acord sobre una metodologia per a seleccionar un programari amb la funcionalitat necessària per a complir els criteris de fiabilitat.

En resum, l'*Audit checklist* proposa un esquema de criteris molt detallat, la qual cosa permet que els centres avaluin la seva actuació només relacionada amb les funcions que hagin acceptat dur a terme. Aquesta categorització és útil, ja que algunes institucions no assumiran la responsabilitat de totes les activitats relacionades amb la preservació sinó que optaran per encarregar determinats serveis a altres organitzacions: empreses comercials, altres institucions amb

Relació amb l'original

Es pot recordar que no sempre es lliurarà la còpia idèntica al document original: per exemple, un document creat en Excel podria guardar-se i lliurar-se a l'usuari en format PDF. L'important és que l'objecte contingui les metadades suficients per a identificar com era l'original i explicar en què consistien els canvis efectuats.

més capacitat o dipòsits especialitzats en la preservació de determinats formats o tipus de recursos. L'externalització d'operacions de preservació es comenta en la pròxima secció.

Adequació d'OAIS

Un informe britànic va analitzar l'adequació d'OAIS com a model per als dipòsits institucionals que no tinguessin com a prioritat la preservació digital. La conclusió va ser que, a causa del nivell de detall, resultaria massa complex per a molts d'ells. La inversió temporal i econòmica necessària per a la implantació d'un dipòsit OAIS seguint els criteris de l'*Audit checklist* representaria una barrera per a determinats centres. L'informe acaba recomanant l'adopció de les línies generals de l'OAIS, però exclouent en principi els criteris detallats. Com a alternativa, se suggereix l'elaboració i ús de pautes, recomanacions i mecanismes d'autoavaluació "light" per a la majoria de dipòsits institucionals a la Gran Bretanya (Allinson, 2005).

3.3.2. Els criteris elaborats per NESTOR

Paral·lelament a l'elaboració i avaluació de l'*Audit checklist*, a Alemanya es va desenvolupar un altre conjunt de criteris per a avaluar la fiabilitat dels dipòsits digitals. Es tracta del *Catalogue of criteria for trusted digital repositories*, elaborat per NESTOR (Network of Expertise in Long-Term Storage of Digital Resources), la versió preliminar del qual en anglès va sortir a les acaballes del 2006. També es preveu la creació eventual d'un sistema de certificació basat en aquests criteris. El catàleg agrupa els criteris en tres àrees: organització, gestió de l'objecte i infraestructura i seguretat.

Cada entrada del catàleg consisteix en quatre camps: text del criteri, descripció general del criteri, exemples, comentaris, notes, etc., i bibliografia relacionada. A continuació s'ofereix l'exemple del criteri relacionat amb la comprovació de la integritat del recurs custodiat, dins de l'àrea "Gestió de l'objecte":

Exemple

Text del criteri

Els dipòsits digitals garanteixen la integritat dels objectes digitals durant totes les fases del procés.

Descripció general

La integritat es refereix al fet que cada un dels objectes digitals es conservi en la seva totalitat i que no hi hagi modificacions indesitjades, tal com es defineix en les normes de preservació. Per a mesurar-ne la integritat es tenen en compte les característiques dels materials digitals que es preserven (cf. 9.2.).

Exemples, comentaris, etc.

Les modificacions indesitjades poden ser causades per errors humans (voluntaris o accidentals), imperfeccions tècniques o danys de la infraestructura tècnica o ser causats per ella.

En els dipòsits digitals s'haurien de prendre precaucions tant organitzacionalment com tècnicament per a assegurar la integritat dels objectes digitals sota la seva custòdia.

Els dipòsits digitals haurien d'utilitzar un sistema de gestió adequat per a preservar la integritat en els processos d'entrada de dades, emmagatzemament i accés. Així mateix, s'haurien de prendre precaucions pel que fa a la integritat del sistema de gestió en si mateix.

Un exemple de modificació voluntària o accidental és la introducció de materials infectats amb virus, l'execució dels quals pot donar lloc a la modificació d'elements del siste-

ma o d'altres elements (per exemple, *scripts* de la base de dades que esborrin objectes o metadades).

Són exemples d'imperficcions tècniques: els programaris defectuosos o incomplets, especialment els utilitzats per a manipulacions complexes (com les migracions), i els suports de memòria obsolets o que no hagin estat emmagatzemats conforme a les especificacions. Generalment, tanmateix, les imperficcions tècniques que fossin previsibles haurien de ser remediades o minorades mitjançant l'aplicació de correccions d'errors apropiades o amb procediments d'identificació d'errors. En alguns casos l'usuari pot seleccionar procediments de correcció d'errors d'alt nivell per a certs tipus de components del sistema (per exemple, mitjançant un grau més gran de redundància). Aquesta possibilitat s'hauria d'usar sempre que fos possible.

L'arxiu de preservació s'haurà de comprovar a intervals regulars per a actualitzar-lo o migrar-lo cap a nous formats. El material s'ha de trobar en formats normalitzats actualitzats, especialment pel que fa a gràfics, pàgines amb imatges i altres tipus d'informació a mesura que vagin sorgint. La informació s'ha de migrar cap a nous formats abans que els actuals corrin el perill de quedar-se obsolets o que siguin impossible de llegir, o quan els nous formats proporcionin millores substancials en les seves característiques, sense que es corri el perill de perdre'n el contingut.

Bibliografia

[ISO 15489-1, Information and documentation, Records Management, 2001]

[Network Working Group, Shirey R: *Internet Security Glossary*, 2000]

[ISO/IEC 15408-x:2005 Information technology -- Security techniques -- Evaluation criteria for IT security, 2005]

[Interpares: Ergebnisse des InterPares-Projekts, 2006]

Aquest criteri és comparable al següent de l'*Audit checklist*: Funcions, processos i procediments > Planificació i estratègies de la preservació > "El dipòsit comprova contínuament la integritat del recurs custodiat".

3.4. Externalització de serveis de preservació

Tal com s'ha afirmat, el col·lectiu bibliotecari accepta com un fet la impossibilitat que qualsevol institució, independentment de la grandària, prestigi o recursos disponibles, pugui fer front sola a totes les exigències de la preservació digital.

L'aplicació del model OAIS per a biblioteques, desenvolupada per l'RLG i l'OCLC, reconeix que hi haurà institucions que encarregaran una part de la seva responsabilitat a d'altres que tinguin més coneixement o capacitat per a fer algunes de les operacions necessàries. De fet, aquesta relació és una de les que es basarà en la fiabilitat de l'organització externa i fa evident la necessitat de l'eventual sistema de certificació, ja comentat, perquè les institucions puguin avaluar els serveis externs oferts. A més del nivell de servei ofert, també caldrà explorar les possibles relacions entre les institucions que custodien recursos digitals i els proveïdors de serveis perquè es pugui definir clarament el paper i les responsabilitats de cada un, projecte per projecte (Hockx-Yu, 2006).

En aquesta secció s'exposen alguns exemples de serveis i projectes relacionats amb l'externalització de responsabilitats identificades pel model OAIS.

Hi ha diversos models de servei contractat, però el més freqüent és el que inclou la gestió tècnica dels arxius que pot, o no, incloure l'accés als materials per part dels usuaris. És evident que el sector informàtic és expert en la gestió de dades digitals que s'han d'incloure: emmagatzemament, còpies de seguretat, protecció de dades, migracions, control d'accés. I moltes institucions contracten aquest nivell de gestió, encara que precisament poden faltar els aspectes més directament relacionats amb la preservació a llarg termini, com ara la creació i gestió de les metadades necessàries per a la recuperació i ús de les dades, la vigilància del contingut per a assegurar-ne la llegibilitat i usabilitat en el futur i la garantia de la sostenibilitat dels serveis en el temps.

Molts experts qüestionen la fiabilitat a llarg termini dels serveis comercials i, per això, fomenten l'establiment de centres sense finalitat de lucre, especialitzats en la preservació que prestaran serveis a d'altres. Alguns exemples de dipòsits fiables els serveis dels quals són subcontractats per d'altres:

- Digital Preservation Repository de la California Digital Library (CDL)
- European Archive
- UK Data Archive

El cas del dipòsit de la CDL, iniciat el 2005, és especialment destacable, ja que ofereix un conjunt de funcions i eines que permeten una contractació de serveis *a la carta*: en paraules seves, un model de preservació *self service* que permet una selecció per part de les institucions client que s'adapta a les seves necessitats i capacitat, millor que els serveis fixos. Un avantatge d'un sistema flexible com aquest és que s'incorporen al conjunt d'eines disponibles les aplicacions que desenvolupin els mateixos clients.

Algunes biblioteques nacionals comencen a garantir la preservació a llarg termini a recursos digitals que no es consideren necessàriament com a part del patrimoni bibliogràfic, tradicionalment recollit en el dipòsit legal. Com a exemple, la biblioteca nacional holandesa admet els materials següents que anteriorment, en un format analògic, no s'incorporaven al fons mitjançant el dipòsit legal:

- **Recursos d'investigació de les universitats nacionals⁴**: les universitats holandeses envien còpies dels treballs d'investigació i tesi, entre altres a la KB; encara que les universitats també arxivin i gestionin aquest material al seu propi dipòsit, la responsabilitat de la preservació a llarg termini l'assumeix la biblioteca nacional.
- **Revistes científiques en format digital d'editorials no holandeses⁵**: la funció de la KB en aquests casos no es deu necessàriament a la incapacitat de les editorials de gestionar els seus propis recursos, sinó al fet que les biblioteques d'investigació subscriptores no les considerin prou fiables. Per

⁽⁴⁾DARENET, <http://www.darenet.nl/en/page/language.view/repositories>

⁽⁵⁾E-DEPOT, <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

això insisteixen que es dipositin en un dipòsit segur per a poder garantir que siguin accessibles i utilitzables perpètuament.

En els exemples anteriors, la institució externa s'encarrega de determinades operacions relacionades amb la preservació digital i, en alguns casos, de totes. Però paral·lelament a la creació d'aquest model de servei, també s'observa una altra tendència en la qual la responsabilitat de les operacions es comparteix de manera més equitativa. Dos projectes britànics que representen el model de col·laboració són SHERPA DP i UK Web Archive. En el primer cas, es tracta d'un projecte britànic coordinat per l'Arts and Humanities Data Service (AHDS), la finalitat del qual és desenvolupar una infraestructura per a la preservació compartida, basada en el model OAIS. Els participants s'encarreguen de diferents funcions, amb l'avantatge de poder compartir entre tots els coneixements i experiències que aporta cada un.

El segon cas és una iniciativa, també britànica, iniciada el 2004 sota la coordinació del UK Web Archiving Consortium, en el qual participen, entre altres socis, les biblioteques nacionals britànica, escocesa i gal·lesa i l'arxiu nacional britànic. En aquest model consorciat, cada membre s'encarrega de la selecció de material web del seu propi interès i segons criteris establerts per cada un. L'emmagatzemament, gestió i altres operacions administratives es fan coordinadament, la qual cosa dona com a resultat un estalvi econòmic per a tots (Bailey i Thompson, 2006).

Sigui quin sigui el model triat –centralitzat o descentralitzat–, les iniciatives de responsabilitats de preservació compartides sembla que són el camí més encertat per a reduir al mínim la duplicació d'esforços i així maximitzar el resultat de la inversió feta pels diferents actors implicats.

Orientar les institucions

Un dels primers intents d'orientar les institucions que custodien recursos digitals és el fulllet sobre la contractació de serveis redactat per la Digital Preservation Coalition el 2004: *Contracting out for digital preservation services*. En un to despreocupat, pensat per a introduir el tema a persones sense coneixements previs, comença per tranquil·litzar els responsables de dipòsits que no puguin assumir les operacions tècniques relacionades amb la preservació: "La majoria de nosaltres tampoc no ho podem fer, igual que no podem fer el manteniment dels nostres propis cotxes o arreglar els nostres televisors...".

El text identifica diversos tipus de proveïdors de serveis externs: empreses comercials; institucions homòlogues (altres biblioteques, arxius o museus) amb més capacitat; entitats públiques amb la capacitat tècnica i un interès institucional a protegir les dades (universitat, arxiu nacional). El grau de control assumit pel proveïdor extern variarà segons els serveis que ofereixi i la fiabilitat que doni a la institució. S'aconsella a la institució sobre els aspectes més rellevants que cal concretar abans de firmar un acord per a la gestió del material, per exemple, els relacionats amb:

- **Els recursos digitals:** tipus, format, quantitat, titularitat;
- **Els serveis contractats i la gestió dels recursos:** tipus d'accés i usos previstos, restriccions a l'accés o ús, confidencialitat, seguretat dels recursos, protecció de les dades;
- **El seguiment del compliment dels serveis contractats:** indicadors del bon funcionament, tipus d'informes i estadístiques sobre l'estat dels recursos i la freqüència del seu lliurament a la institució.

Com s'ha comentat en la secció anterior, s'intenten establir sistemes d'avaluació i certificació de dipòsits digitals com, per exemple, el servei programat pel DCC a la Gran Bretanya. Quan finalment es materialitzin aquests sistemes, es facilitarà molt la tasca de les institucions menys capacitades per a escollir proveïdors amb més garantia de la seva fiabilitat.

3.5. Els dipòsits institucionals i la preservació

En els últims anys s'ha detectat un creixement impressionant del nombre de dipòsits institucionals (*institutional repositories*) implantats en moltes organitzacions –sobretot en universitats. La seva finalitat és emmagatzemar i gestionar els recursos digitals creats per persones o dependències dins d'una mateixa institució, ja que, cada vegada més, aquests recursos es consideren com a béns intel·lectuals (*intellectual assets*) institucionals que mereixen que se'ls custodii.

Però els objectius concrets d'aquests dipòsits poden variar, per exemple:

- la gestió de la gran varietat de recursos produïts –amb finalitats docents, científiques i/o administratives– per a facilitar:
 - la generació d'estadístiques;
 - la reutilització (*repurposing*) d'elements dels recursos per part d'altres: aquest objectiu se sol centrar a augmentar l'aprofitament dels recursos existents, sobretot per a material docent;
 - el compliment de les exigències de les agències d'avaluació i finançament que exigeixen que les universitats demostrin el nivell de rendiment i els resultats concrets produïts.
- la disponibilitat i fàcil accés de material científic produït per investigadors de la institució, d'acord amb el moviment Open Access;
- la preservació de material digital produït en qualsevol de les dependències per a la seva localització i utilització en el futur.

A causa de les diferents finalitats dels dipòsits ja instal·lats, s'observa un alt grau de divergència entre ells quant a factors com els següents:

- el tipus de material admès;
- l'autoria dels recursos i l'autorització per a fer el dipòsit;
- els formats acceptats i mantinguts;
- els processos i procediments efectuats;
- el control de l'accés, i
- la durada de la gestió.

La varietat de possibles motius i, com a conseqüència, la multitud de perfils i aplicacions existents creen, sovint, un clima de confusió sobre quina és la funció principal del dipòsit institucional, fins i tot entre els més directament relacionats amb el seu funcionament.

Sigui quina sigui la finalitat, moltes de les actuacions dutes a terme per aquests dipòsits s'encavalquen en certa manera: gestió física de les dades, descripció del contingut mitjançant metadades, control de seguretat. Tanmateix, les activitats no són sempre complementàries:

- molt de material d'interès administratiu o custodiat per a la seva preservació no es deixarà en règim d'accés obert;
- molt de material d'investigació o administratiu tindrà un cicle vital de termini fix, sense la necessitat de preservar-lo a llarg termini.

A la pràctica, a escala internacional es constata que encara són pocs els dipòsits institucionals que tenen la preservació com a prioritat; molts, fins i tot, ni tan sols reconeixen la problemàtica que representa. L'estudi britànic *Digital repositories roadmap* va observar fins i tot el poc consens sobre el paper que tindran aquests dipòsits quant a la preservació en el futur. El debat sembla centrar-se, segons aquest informe, en qui assumirà les responsabilitats de la preservació dels recursos digitals britànics: si cada institució serà responsable de la seva pròpia producció digital o si es crearan serveis especialitzats nacionals que s'encarregaran d'aquestes operacions per a assegurar-ne l'accessibilitat i usabilitat en el futur (Heery i Powell, 2006).

És important destacar que encara que el dipòsit digital sigui un component bàsic per a la preservació a llarg termini dels recursos digitals, la seva implantació en una institució no és mostra ni garantia que els recursos custodiats es preservaran a llarg termini.

Bibliografia

Abrams, Stephen L. (2005). "Establishing a global digital format registry". *Library Trends* (vol. 1, núm. 54, pàg. 125-143).

Abrams, Stephen L. (2004). "The role of format in digital preservation". *Vine: The Journal of Information and Knowledge Management Systems* (vol. 2, núm. 34, pàg. 29-55).

Abrams, Stephen L.; Flecker, Dale (2005). *Proposal for a global digital format registry*. <<http://hul.harvard.edu/gdfr/documents/Proposal-2005-09-29.doc>>

Allinson, Julie (2005). "OAIS as a reference model for repositories: an evaluation". Revised draft 0.4 UKOLN, 2005. <<http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/>>

An Audit checklist for the certification of trusted digital repositories; draft for public comment (2005). RLG; NARA. <<http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>>

Authenticity in a digital environment (2000). Council on Library and Information Resources. <<http://www.clir.org/pubs/reports/pub92/contents.html>>

Authenticity of Electronic Records. <www.ica.org/biblio/Study13_2Erev.pdf>

Availability & preservation: long term availability & preservation of digital information; AIIM industry white paper on records, document and enterprise content management for the public sector (2002). Berkshire, Regne Unit: AIIM Europe.

Bailey, Steve; Thompson, Dave (2006). "UKWAC: building the UK's first public web archive". *D-Lib Magazine* (v. 12:1). <<http://www.dlib.org/dlib/january06/thompson/01thompson.html>>

Bates, Melanie i altres (2006). *Digital lifecycles and file types: final report*. JISC; Loughborough University. <<https://magpie.lboro.ac.uk/dspace/bitstream/2134/1793/1/DigitalLifecyclesV2.pdf>>

Beachum, Frank. "Archivists warn: don't depend on digital tape". <http://www.minidisc.org/dat_archiving.html>

Bearman, David; Trant, Jennifer (1998, juny). "Authenticity of digital resources: towards a statement of requirements in the research process". *D-Lib Magazine*. <<http://www.dlib.org/dlib/june98/06bearman.html>>

California Digital Library (2005). *Preserving digital materials: creating an Open Archival Information System (OAIS) conformant preservation repository for multi-institutional use*. <libraries.universityofcalifornia.edu/sopag/dpac/DPACFinalReport.pdf>

Candás Romero, Jorge (2006, març-abril). "El papel de los metadatos en la preservación digital". *El Profesional de la Información* (vol. 15, núm. 2, pàg. 126-136).

Cedars guide to digital preservation strategies (2002). <www.leeds.ac.uk/cedars/guideto/dpstrategies>

Conway, Paul (1996). *Preservation in the digital world*. <<http://www.clir.org/pubs/reports/conway2/>>

Data dictionary for preservation metadata: final report of the PREMIS Working Group. OCLC; RLG, 2005. <<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>>

Defining authenticity in the digital environment. <<http://www.cni.org/tfms/2000a.spring/handout/Smith-Authenticity2000Stf.pdf>>

Digital Preservation Coalition (2004). *Contracting out for digital preservation services: information leaflet and checklist*. <<http://www.dpconline.org/docs/guides/outsourcing.pdf>>

Digital Preservation Coalition (2005). *Directory of digital preservation repositories and services*. <<http://www.dpconline.org/docs/guides/directory.pdf>>

Directrices para la preservación del patrimonio digital (preparat per la Biblioteca Nacional d' Austràlia) (2003). UNESCO.

<<http://www.unesdoc.unesco.org/images/0013/001300/130071s.pdf>>

Establishing a central depository for preserving digital image collections Part 1: Responsibilities of transferee (2001). Cornell, Nova York: Cornell University Library.

<<http://www.library.cornell.edu/imls/image%20deposit%20guidelines.pdf>>

File formats typology and registries for digital preservation (2004). DELOS Project.

<[www.dpc.delos.info/private/output/DELOS_WP6_d631_finalv2\(5\)_urbino.pdf](http://www.dpc.delos.info/private/output/DELOS_WP6_d631_finalv2(5)_urbino.pdf)>

Geser, Guntram; Pereira, John (eds.) (2004). "The future digital heritage space : an expedition report". *DigiCult Thematic* (núm. 7).

<http://www.digicult.info/downloads/dc_thematic_issue7.pdf>

Gladney, H. M. [Missatge de correu electrònic enviat a la llista de distribució JISC-REPOSITORIES, el 12 d'abril de 2006.]

Granger, Stewart; Russell, Kelly; Weinberger, Ellis (2000). "Cost elements of digital preservation" (Ver. 4).

<<http://www.leeds.ac.uk/cedars/colman/costElementsOfDP.doc>>

Handbook for digital projects: a management tool for preservation and access. Northeast Document Conservation Center (2000).

<http://www.nedcc.org/oldnedccsite/digital/dman2.pdf>

Hedstrom, Margaret; Montgomery, Sheon (1998). *Digital preservation needs and requirements in RLG member institutions*. Mountain View, Califòrnia: Research Library Group.

<<http://www.rlg.org/preserv/digpres.pdf>>

Heery, Rachel; Powell, Andy (2006). *Digital repositories roadmap: looking ahead (version 15 UKOLN)*.

<http://www.jisc.ac.uk/uploaded_documents/rep-roadmap-v15.doc>

Hockx-Yu, Helen (2006). "Digital preservation in the context of institutional repositories". *Program: electronic library and information systems* (vol. 3, núm. 40, pàg. 232-243).

Hodge, Gail M. (2000). "Best practices for digital archiving: an information life cycle approach". *D-Lib Magazine* (vol. 1, núm. 6).

<<http://www.dlib.org/dlib/january00/01hodge.html>>

IFLA (2000). *Principios para el cuidado y manejo de material de bibliotecas* (traduït al castellà). Santiago de Xile: DIBAM.

<<http://www.ifla.org/V1/4/news/pchlm-s.pdf>>

Jones, Maggie; Beagrie, Neil (2001). *Preservation management of digital materials: a handbook*. Londres: British Library.

<<http://www.dpconline.org/graphics/intro/index.html>>

Jorum preservation watch report (2006). JISC.

<http://www.jorum.ac.uk/docs/pdf/Digital_Preservation_Report.pdf>

Kaczmarek, Laura, i altres (2006, desembre) "Using the Audit checklist for the certification of a trusted digital repository as a framework for evaluating repository software applications: a progress report". *D-Lib Magazine* (12:12).

<<http://www.dlib.org/dlib/december06/kaczmarek/12kaczmarek.html>>

Kenney, Anne i altres (2003). *Digital preservation management: implementing short-term strategies for long-term problems*. Tutorial.

<www.library.cornell.edu/iris/tutorial/dpm/>

Lamonica, Martin. "Opendocument discussion veers toward consumers". *Open source and standards* (15 de desembre de 2005).

<http://news.com.com/2061-10795_3-5996639.html?part=rss&tag=5996639&subj=news>

Ministerio de las Administraciones Públicas (2004). *Aplicaciones utilizadas para el ejercicio de Potestades: Criterios de conservación*.

<<http://www.csi.map.es/csi/criterios/pdf/conservacion.pdf>>

NESTOR (2006). *Catalogue of criteria for trusted digital repositories. Version 1*.

<<http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf>>

Overview of the UC Libraries Digital Preservation Repository (DPR). Califòrnia Digital Library.

<<http://www.cdlib.org/inside/projects/preservation/dpr/DPRoverview.pdf>>

Pathfinder Research on Web-based Repositories Final Report (2004). Bristol, Regne Unit: Mark Ware Consulting. [Publisher and Library/Learning Solutions]

<[http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/\\$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf](http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf)>

Preserving digital information (1996). Mountain View, Califòrnia: Research Library Group.

<<http://www.rlg.org/ArchTF/tfadi.index.htm>>

Reference model for an open archival information system (OAIS) (2002). Blue Book. Consultative Committee for Space Data Systems.

<<http://public.ccsds.org/publications/archive/650x0b1.pdf>>

Report to ICABS on guidance for digital preservation: report on a survey of sources (2005). National Library of Australia.

<<http://www.nla.gov.au/nla/staffpaper/2005/documents/webb1.pdf>>

Ross, Seamus; McHugh, Andrew (2006, juliol i agost). "The role of evidence in establishing trust in repositories" *D-Lib Magazine* (núm. 12, pàg. 7/8).

<<http://www.dlib.org/dlib/july06/ross/07ross.html>>

Rothenberg, Jeff (2000). *An experiment in using emulation to preserve digital publications* (pàg. 69). La Haia: Koninklijke Bibliotheek.

<<http://nedlib.kb.nl/results/emulationpreservationreport.pdf>>

Rothenberg, Jeff (1999). *Avoiding technological quicksand: finding a viable technical foundation for digital preservation*. Washington: CLIR.

<<http://www.clir.org/PUBS/reports/rothenberg/contents.html>>

Rusbridge, Adam (2003). "Migration on request; 4th year project report: computer science". University of Edinburgh, Division of Informatics.

Rusbridge, Chris (2006, 8 de febrer). "Excuse me... Some Digital Preservation Fallacies?". *Ariadne* (núm. 46).

<<http://www.ariadne.ac.uk/issue46/rusbridge/intro.html>>

Salvador Benítez, Antonia; Ruíz Rodríguez, Antonio (2005). "Metadatos para la preservación de colecciones digitales". *Cuadernos de Documentación Multimedia* (v. 16).

<<http://multidoc.rediris.es/cdm/index.php>>

Shirky, Clay (2005). "AIHT: conceptual issues from practical tests". *D-Lib Magazine* (v. 11, pàg. 12).

<<http://www.dlib.org/dlib/december05/shirky/12shirky.html>>

Smith, Abby (2000). *Authenticity in a digital environment*. Washington, DC: CLIR.

<www.clir.org/PUBS/reports/pub92/pub92.pdf>

Smith, Abby (1998, maig i juny). "Preservation in the future tense". *CLIR Issues* (núm. 3).

<<http://www.clir.org/pubs/issues/issues03.html#preserve>>

Swan, Alma; Brown, Sheridan (2005). *Open access self archiving: an author study*. Truro, Regne Unit: Key Perspectives.

<<http://eprints.ecs.soton.ac.uk/10999/>>

Trusted digital repositories: attributes and responsibilities.

<<http://www.rlg.org/legacy/longterm/repositories.pdf>>

U. K. Data Archive. *Recommendations*. <<http://www.data-archive.ac.uk>>

Verheul, Ingeborg (2006). *Networking for digital preservation: current practice in 15 national libraries*. Munic: K. G. Saur ("IFLA Publications", 119).

<<http://www.ifla.org/V/pr/saur119.htm>>

Waller, Martin; Sharpe, Robert (2006). *Mind the gap: assessing digital preservation needs in the UK*. Digital Preservation Coalition.

<<http://www.dpconline.org/docs/reports/uknamindthegap.pdf>>

Wheatley, Paul (2004). *Institutional repositories in the context of digital preservation*. Digital Preservation Coalition (DPC Technology Watch Series Report 04-02).

<http://www.eprints.rclis.org/archive/00007351/01/DPinIRs_Final.pdf>