Preservació digital

Daniel Martínez Armada
Enginyeria d'informàtica

José Manual Castillo Pedrosa

08/06/2016

Aquesta obra està subjecta a una llicència de Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons

FITXA DEL PROJECTE FINAL

Títol del projecte:	*Preservació digital*
Nom de l'autor:	*Daniel Martínez Armada*
Nom del consultor:	*José Manual Castillo Pedrosa*
Data de lliurament (mm/aaaa):	*06/2016*
Àrea del Projecte Final:	*Administració de Xarxes i Sistemes Operatius*
Titulació:	*Enginyeria d'informàtica*

Resum del Projecte (màxim 250 paraules):

Anàlisi i estudi de un projecte de preservació de dades digitals. Anàlisi dels elements que ho conformen, dels perills que amenaces actualment a les dades i als seus suports físics i presentació de solucions mitjançant tècniques, eines i polítiques destinades a preservar les dades més enllà de la vida útil d'un format en concret o el suport que les conté.

Estudi d'un cas pràctic: biblioteca universitària. Elecció de les eines i tècniques, així com de la infraestructura tècnica i del programari més adient per a assolir la tasca de la preservació digital del fons bibliogràfic de la biblioteca, seguint el model de referència OAIS.

Abstract (in English, 250 words or less):

Analysis and study of a digital preservation project. Analysis of the elements that form it, dangers threatening our data and hardware, explaining solutions and techniques, tools and policies that protect data beyond the span life of a particular format or digital device containing them.

Practical Case Study: University library. Choice of tools and techniques, as well as technical infrastructure and software best fit to achieve the task of preserving digital library collection, following the OAIS reference model.

Paraules clau (entre 4 i 8):

Preservació digital, obsolescència, digitalització, repositori, OAIS

Índex

1. Introducció 1
1.1. Context i justificació del Projecte 1
1.2. Objectius del Projecte 1
1.3. Enfocament i mètode seguit 1
1.4. Planificació del Projecte 1
1.5. Breu sumari de productes obtinguts 2
1.6. Breu descripció dels altres capítols de la memòria 2
2. Anàlisi teòric 3
2.1. La preservació digital. Definició i conceptes 4
2.2. Aspectes tècnics 9
2.3. Aspectes no tècnics. Gestió de la preservació. 15
2.4. Tècniques i estratègies de preservació. 18
2.5. Eines per a la preservació 23
2.6. Model OAIS 29
3. Cas d'estudi d'una biblioteca universitària 32
3.1. Planificació de la preservació 34
3.2. Administració i gestió del dipòsit digital 46
3.3. Ingesta 49
3.4. Magatzem de dades 54
3.5. Gestió de les dades 59
3.6. Accés 62
Annex. Motius per dur a terme un refresc de suports. 67
4. Conclusions 74
5. Glossari 76
6. Bibliografia 77

Llista de figures

Diagrama 1. Planificació inicial del projecte..... 1
Taula 1. Factors de degradació de suports 12
Diagrama 2. Model de dades de PREMIS.....27
Diagrama 3. Model OAIS 29
Diagrama 4. Model d'informació OAIS expandit 33
Taula 2. Classificació de formats segons la seva preferència al nostre sistema..... 35
Diagrama 5. Gantt per a la planificació temporal del projecte de preservació..... 37
Taula 3. Anàlisi de riscos 40
Imatge 1. Captura de pantalla del software Rosetta..... 42
Imatge 2. Visor de metadades de XENA 42
Imatge 3. Rack de 4 discs SATA.....54
Imatge 4. Mirroring de RAID1. 55
Imatge 5. Cintes magnètiques LTO-7 57
Imatge 6. Paràmetres SMART 58
Imatge 7. Visor del software de preservació XENA.....63
Imatge 8. Varis sistemes operatius executant damunt una distribució Linux..... 65
Imatge 9. Tassa de errades segons temperatura d'un disc dur Hitachi.....68
Imatge 10. Banda antiestàtica per evitar ESD (Electrostatic discharge) 70
Imatge 11. SAI Online..... 70
Imatge 12. Plat afectat per capçal..... 71
Imatge 13. Preu per Gigabyte en dòlars americans.....72
Imatge 14. Comparativa connectors PATA i SATA 73

1. Introducció

1.1. Context i justificació del Projecte

La duració de les dades depèn del format i el suport en el que s'emmagatzemen. Aquests tenen una vida útil limitada, causant un problema a organitzacions i institucions que volen que les seves dades perdurin per molt més temps.

1.2. Objectius del Projecte

Establir polítiques, tècniques i eines orientades a una duració prolongada de les dades, més enllà de la vida d'un determinat format o dispositiu.

1.3. Enfocament i mètode seguit

S'ha fet un estudi teòric del estat actual de les tècniques i eines referides a la preservació digital, analitzant el seu impacte i definint les pautes que han de guiar a un projecte d'aquesta envergadura.

Es proposa un cas pràctic com a subjecte d'estudi i de elecció entre les diferents alternatives de que disposem.

1.4. Planificació del Projecte

Id	Nombre de tarea	Duración	Comienzo
1	Pla de treball	2 sem.	lun 07/03/16
2	**Part teòrica**	**5 sem.**	**lun 21/03/16**
3	La preservació digital. Definició	4 días	lun 21/03/16
4	Aspectes tècnics	5 días	vie 25/03/16
5	Tècniques i estratègies de preservació	4 días	vie 01/04/16
6	Eines per a la preservació	3 días	jue 07/04/16
7	Metadades	2 días	mar 12/04/16
8	Model OAIS	5 días	jue 14/04/16
9	Gestió de recursos	2 días	jue 21/04/16
10	Entrega PAC2	0 días	vie 15/04/16
11	**Part pràctica**	**1 ms**	**lun 25/04/16**
12	Definició del problema	5 días	**lun 25/04/16**
13	Oferir més d'una solució. Comparativa	7 días	lun 02/05/16
14	Elaboració d'un pla	8 días	mié 11/05/16
15	Entrega PAC3	0 días	vie 20/05/16
16	**Conclusions**	**2 sem.**	**lun 23/05/16**
17	Estudi actual dels suports digitals	5 días	**lun 23/05/16**
18	Evolució dels suports digitals	4 días	lun 30/05/16
19	Conclusions	1 día	vie 03/06/16
20	Presentació Power Point	2 sem.	lun 23/05/16
21	Entrega documentació	0 días	mié 08/06/16

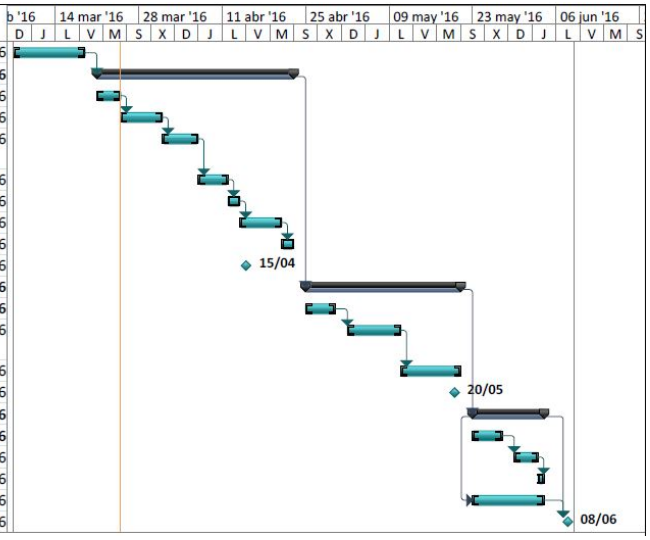


Diagrama 1. Planificació inicial del projecte

1.5. Breu sumari de productes obtinguts

Present document amb anàlisi teòric i cas pràctic (pla de preservació).
Presentació Powe Point resum

1.6. Breu descripció dels altres capítols de la memòria

Aquesta memòria s'estructura en tres capítols, l'anàlisi teòric, el disseny d'un pla de preservació per a una biblioteca universitària, i un petit annex que parla dels motius principals per dur a terme refresc de suports.

2. Anàlisi teòric

Introducció

La conservació i preservació de la cultura i informació obtinguda al llarg dels mil·lennis és fonamental per a la pervivència i millora de la nostra espècie. Durant tota la Historia, l'home ha intentat emmagatzemar i transmetre el seu coneixement a les generacions futures, possibilitant en molts de casos un base per millorar i aprendre noves tecnologies o per entendre com funciona el mon.

Fins la arribada dels ordinadors, tota la informació emmagatzemada per l'home era analògica, essent el paper el medi més comú per plasmar uns coneixements que haurien de perdurar. Al segle XX, amb la arribada de l'era digital, la informació va començar a enregistrar-se a medis electrònics digitals. Les característiques tècniques peculiars dels suports digitals, fan que la manera de preservar informació adquirida amb els suports analògics durant els segles ja no és vàlida, encara que veurem que en podem reutilitzar qualque idea.

Ja ben entrats al s.XXI la informació digital no només és un aspecte important de la nostra economia i societat, és un pilar fonamental, i com a tal, no volem que sigui feble. Amb la arribada de Internet, les publicacions han augmentat exponencialment. Ara és més fàcil que mai escriure i arribar a milers de persones, qualsevol tipus de literatura, des de l'oci fins a les publicacions científiques. La nostra economia és cada vegada més depenent de una informació digital que pareix que ho inunda tot, des de els pressuposts generals d'un Estat fins a les Borses de tot el món.

Per acabar d'afegir més volum de dades digitals a les ja existents, s'ha de parlar dels esforços que grans empreses, universitats i biblioteques estan fent en la digitalització de totes les seves publicacions analògiques. Com a mecanisme per fer-les perdurar més en el temps o simplement per disposar d'una còpia per una millor consulta.

És important establir amb serietat com perdurarem tot aquest ingent volum de dades generat en els darrers 40 anys en aquest frenesí de documentació digital i al que estem tan exposats tant culturalment com econòmicament.

Aquest primer capítol del projecte de final de carrera (PFC) es basa en estudiar la base teòrica de totes les eines i tècniques actuals per a assolir la preservació digital, entesa com un concepte global on hi haurà molts d'actors implicats, no només tècnics, i on la economia hi té molt a dir.

2.1. La preservació digital. Definició i conceptes

Què és la preservació digital?

Definició:
Consisteix en la aplicació de tècniques i metodologies que ens garanteixen que la informació emmagatzemada digitalment en qualsevol tipus de format, programari, màquina o sistema continuï essent accessible en el futur.

La informació digital és, al cap i a la fi, un conjunt de bits agrupats d'un determinat mode (format) que permet la seva posterior lectura per un programari preparat per llegir-la. Aquest programari està desenvolupat per fer-se servir damunt un maquinari específic. Tots 3 conceptes (format, programari, maquinari) tenen una vida útil limitada. El maquinari, per que els components dels quals està fabricat acaben per fallar quan arriben al final de la seva vida útil, o bé per que queda obsolet degut a la aparició de maquinari més modern, amb més prestacions. El programari, per que està desenvolupat per a un maquinari o sistema en concret, fent que si el sistema canvia, s'hagi de substituir per una versió més moderna. Els formats tendeixen a durar més que el maquinari i el programari, però no són eterns. A mesura que canvien les arquitectures de maquinari i el software s'adapta a aquestes, hi surten nous formats que s'adapten millor a les noves formes d'emmagatzemar la informació.

Tota aquesta obsolescència té una ràpida solució, substituir el maquinari o programari vell per un nou. El problema ve quan les dades que volem preservar solen tenir una vida útil que va més enllà de la vida útil de maquinar, programari o formats. Si no establim un pla de preservació digital per a les nostres dades, no serem capaços de tornar-les a llegir una vegada s'hagi canviat de tecnologia. És per això que ha d'existir el concepte de preservació digital i de totes les tècniques associades per a assolir-la.

S'ha de tenir en compte un detall addicional que posa en perill les nostres dades digitals: la seva modificació és relativament senzilla comparada amb les dades analògiques i es pot fer tant deliberadament com de forma accidental. En tots dos casos les dades que volem preservar han canviat i això s'ha de impedir o s'ha d'enregistrar, en cas de ser una modificació desitjada. En aquest cas cobrarà molta importància la seguretat informàtica. També s'han de posar en marxa mecanismes per evitar una corrupció de dades quan aquestes s'estan manipulant, ja que un petit canvi de uns pocs bytes d'un fitxer el pot deixar totalment inservible.

Què volem preservar?

És important, abans de continuar, definir al nostre pla de preservació quines seran les dades que voldrem preservar. Està clar que no voldrem preservar-ho tot especialment per el alt cost econòmic que això podria tenir damunt la nostra organització. Definir un filtre de dades per a preservar ens ajudarà a reduir costos i a ser més eficients en la nostra cerca d'informació llevant tot allò innecessari.

Uns bons criteris per definir quin material es vol preservar són:

- Obligatorietat legal. Hi ha material que s'ha de preservar durant un mínim de temps establert per la llei.
- Interès especial per a la institució. Aquells materials que són d'especial interès per a la nostra organització, ja sigui per interès econòmic o com referència històrica.
- Us actual i potencial. Molt relacionat amb l'anterior punt. Si es preveu que aquest material el farem servir en un futur tant pròxim com llunyà.
- Valor artístic. Cas especial de biblioteques i fons documentals.
- Disponibilitat a altres magatzems de dades. S'ha d'estudiar si aquest material ja està cobert per un pla de preservació a altres institucions. Això no vol dir que l'haguem d'excloure del nostre pla obligatòriament.
- El cost de preservar el material. Als documents digitals, el cost està molt relacionat amb el pes dels fitxers, ja que un dels majors costos són els suports físics on emmagatzemem la informació, encara que no és l'únic.
- La disponibilitat de personal i instal·lacions per a realitzar el tractament de les dades. Íntimament lligat amb el punt anterior, personal i instal·lacions són dos punts organitzatius claus per a dur a terme el nostre pla de preservació.
- El valor del document.
- Permisos d'accés al fitxer, límits legals a la preservació de determinades dades. Propietat intel·lectual.

Tots aquests criteris, i no són pocs, s'han de tenir en compte alhora d'avaluar si un fitxer s'ha de preservar o no. No és una tasca trivial i el pla de preservació haurà d'especificar i concretar molt clarament quin són els criteris d'inclusió del material a preservar.

Quan hem de començar el tractament?

En aquest punt estudiarem el moment en el que començar a fer el tractament de preservació de les dades digitals des de diferents punts de vista.

Actuació reactiva VS Actuació proactiva.

Actuació reactiva és actuar quan s'ha detectat qualque problema a les nostres dades. Hem d'establir mecanismes per detectar aquests problemes i tractar de recuperar les dades. Aquesta tècnica no és gaire recomanada per al tractament de dades digitals, ja que molt probablement, quan s'ha detectat una errada a les dades ja és massa tard per recuperar-les.

Actuació proactiva és actuar abans de detectar problemes a les dades, establint un pla de prevenció. És el mode recomanat d'actuació damunt dades digitals degut a la seva naturalesa. En comparació amb les dades analògiques, els suports digitals tenen una vida útil molt més curta. Si comparem la vida útil d'un llibre amb la d'un disc dur, tot dos en condicions ideals de conservació, veurem que un llibre pot durar segles i un disc dur probablement no duri ni mig.

És per aquest motiu i molts altres que es recomana actuar damunt la informació digital lo més aviat possible. En el seu moment de creació o obtenció. En el moment que el repositori de dades rep el fitxer digital, ja s'hauria de posar damunt el control de preservació si és d'un tipus afectat (definit al punt anterior). Això requereix de una gran implicació dels creadors de continguts amb l'administrador del repositori i fa aquesta tasca complicada.

Actuació esporàdica VS Actuació continuada

Les actuacions esporàdiques són habituals als plans de preservació de documentació analògica, però no son recomanats per als documents digitals. Això és per que els canvis en la tecnologia, en la vida útil dels suports i la facilitat de modificació de les dades, fan que s'hagi d'establir una actuació continuada damunt els documents digitals.

Omissió VS Actuació

Als tractament d'informació analògica és comú tancar els documents a un lloc segur i amb bones condicions d'emmagatzemament i no tornar a accedir-hi durant anys o inclús dècades. És lo que entenem per omissió. L'accés a aquestes dades analògiques pot destruir-les.

En canvi, per tot lo que hem parlat abans de la vida útil dels suports físics i també del software, no és bona idea emmagatzemar un fitxer digital tal i com el varem trobar al seu dia i no tornar a accedir-hi durant dècades. Si férem això, probablement quan volguéssim llegir el fitxer, no tendríem els mètodes per fer-ho, ja que la tecnologia haurà canviat tant, que no comptaríem amb el programari ni el maquinari per obrir

aquest fitxer. Per això i com veurem més endavant comptarem amb tècniques que ens ajudaran a preservar el document digital al llarg del temps, com ara el refresc de suports.

Qui s'ha d'implicar al tractament?

Persones implicades

El personal que fins ara ha fet feina amb la preservació dels recursos analògics tradicionals haurà d'adaptar-se a les noves metodologies que implica el canvi a digital. Tot el coneixement adquirit fins ara amb la preservació analògica és pot aprofitar, però s'ha d'incorporar tota una sèrie de nous coneixements per adaptar-se al nou model digital, com ara:

- Coneixements d'informàtica
- Qüestions legals relacionades amb la informació digital
- Estar preparats per a tots els canvis que implica la informació digital: canvi de formats, de tecnologia o de legislació…
- Nous canals de comunicació amb les noves professions derivades del món digital:
    - Administradors institucionals
    - Autors i creadors dels recursos digitals
    - Altres contactes institucionals, com ara servei informàtic o assessoria jurídica.
    - Altres contactes externs, com ara editors, organitzacions internacionals, biblioteques nacionals, administracions públiques… etc.

Institucions implicades

Actualment es considera el factor institucional, i no el tècnic, com el més dificultós per al avanç de la preservació digital. La implicació institucional és bàsica per a garantir l'èxit del nostre pla de preservació. Aquesta implicació s'ha de matisar en dos conceptes:

- Viabilitat institucional: s'han de posar els mitjans tècnics, econòmics i materials per a assumir la preservació digital.
- Viabilitat econòmica: s'ha d'assegurar un nivell econòmic i mantenir-lo durant el futur.

Per a assolir aquesta implicació institucional i que sigui viable en els dos conceptes esmentats, s'ha de reconèixer aquests conceptes de forma explícita a la formalització dels nous processos de preservació.

Preservació digital VS Preservació mitjançant digitalització

És important destacar la diferència entre els conceptes de “Preservació digital” i “Preservació mitjançant digitalització”.

La Preservació mitjançant digitalització (o simplement “digitalització”), moltes vegades confosa amb la Preservació digital, és la tècnica mitjançant es vol preservar un recurs analògic fent una còpia digital, normalment amb mètodes de reconeixement òptic (escàners). Aquest tipus de preservació va ser molt comú als anys 90 quan, gràcies als avanços en emmagatzemament digital, es va començar a digitalitzar gran quantitat de documentació.

Els avantatges d’aquesta tècnica són molts, com ara:

- Evitar accessos al document original, ajudant a la seva preservació
- Obtenir una còpia perfectament vàlida i fiable
- Facilitat en la cerca del document
- Accés en línia, sense límits geogràfics
- Accés simultani de més d’un usuari
- Millor presentació
- Fàcil reproducció sense tornar a accedir al original

Però com hem parlat abans, els documents digitals s’enfronten també als seus propis problemes, així que aquests nous documents digitals s’han d’incloure en el pla de preservació digital del fons per no perdre aquesta informació. Recordem que digitalització no és Preservació Digital.

2.2. Aspectes tècnics

En aquest apartat veurem més en profunditat els aspectes tècnics relacionats amb la Preservació Digital que fins ara hem anomenat. Els dividirem en diferents categories:

- Els relacionats amb la naturalesa pròpia del recurs digital.
- Les possibles amenaces tecnològiques a les quals ens enfrontem.
- Els elements a preservar de la informació digital
- Anàlisi de riscos

Naturalesa del recurs digital

En aquest punt en parlarem dels aspectes tècnics que són propis dels recursos digitals i que, per la seva naturalesa, no existien amb la preservació de dades analògiques. Aquests són:

- Codificació de dades. Formats.
- Selecció del recurs
- Mètode emprat per a preservació

Codificació de dades. Formats.

Es el aspecte tècnic més important d'un recurs digital. Com està codificat. Quin és el seu format.

Definim el format d'un fitxer digital com la codificació estructurada de les dades per a la seva gestió i manipulació mitjançant un programa informàtic. De formats n'hi ha de dos tipus: estandaritzats (hi ha una amplia gama de programari capaç d'accedir-hi) i propietaris (formats tancats que necessiten d'un programari específic per ser tractat).

La quantitat de formats existents actualment és molt gran i continua en creixement degut a la creixent quantitat d'aplicacions i dispositius que apareixen en el mercat cada any. Per això, el format del recurs digital a ser preservat és un element clau que ens pot condicionar la resta de punts del recurs digital (mètode emprat, facilitat per recuperar el document…). El Global Digital Format Registry (GFDR), és un projecte que es basa en emmagatzemar i suministrar als dipòsits digitals informació fiable dels formats que en gestionen. La seva informació és molt valuable per a tots els projectes de preservació digital de grans fons ja que comptaran amb molts formats diferents.

Podem classificar els formats segons la seva tipologia. Hi ha vàries classificacions, com per exemple aquesta de la National Library of Australia:

- General
- Documents
- Audiovisuals
- Art digital
- Literatura electrònica
- Jocs, realitat virtual i software
- Manuscrits digitals
- Imatges
- Miscelània
- Recursos web
- Materials digitals en suports físics
- Dades científiques i d'investigació
- Dades geo-espacials
- Notícies o informació d'actualitat

Segons els seus autors, aquesta llista no és exhaustiva ni definitiva, i un fitxer pot pertànyer a més d'una tipologia. Altres autors han classificat els formats segons la seva facilitat d'us i gestió. Per a cada tipus de format es determinen quins són els formats preferits, quins no ho són tant, i quins no són desitjables.

Selecció del recurs

El format triat per a la preservació ens pot condicionar la selecció d'un determinat recurs, en base a si el format és obert o propietari, la qual cosa redundarà en els costos del programari per obrir el fitxer, i també en la facilitat d'ús i accés en el futur.

Mètode emprat per a la preservació
La selecció d'un determinat format pot condicionar el mètode emprat per a la serva preservació, per exemple per la seva complexitat.

Amenaces

Degut a les característiques pròpies dels recursos digitals com ara la seva dependència dels elements informàtics o la seva volatilitat, surten tota una sèrie d'amenaces.

Obsolescència tecnològica

Es refereix al concepte de que un determinat dispositiu o programari, encara que continua funcionant, deixa de ser útil per que el seu entorn ha canviat degut a la ràpida evolució dels sistemes informàtics. Un exemple molt clar d'obsolescència hardware són els disquettes de 3 ½ que, encara que avui en dia podrien funcionar, ja no hi ha pràcticament cap ordinador que incorpori un lector per a ells.

L'entorn hardware

Concepte molt semblant a l'anterior però més global, ja que actualment els ordinadors estan deixant passar a altres dispositius com ara smartphones, tablets per a la generació de continguts.

L'entorn software

Anàlogament amb el cas anterior, el programari també té una ràpida evolució i a més, una clara dependència del maquinari per al qual va estar compilat, lo que el fa encara més vulnerable a quedar obsolet.

Suports físics

Menció especial a la obsolescència tecnològica li dedicarem als suports físics, ja que són el lloc físic on emmagatzemem el fitxer. La primera passa per recuperar un fitxer que estigui sota el nostre pla de preservació és recuperar on està emmagatzemat i llegir aquest suport. El problema ve quan no en tenim un lector apropiat per que el suport ha quedat obsolet i ja no en fabriquen més.

En els darrers anys hem vist desaparèixer gran quantitat de tipus de suports físics digitals, des de els disquettes de 5 ¼, 3 ½, ZIP, mini Disc, UMD... i tants altres formats minoritaris en favor d'altres cada vegada més moderns, amb més capacitat i menor preu per bit. Actualment, a part dels discs durs, les opcions són els discs òptics (CD, DVD, BluRay) però és d'esperar que tard o d'hora aquests formats desapareguin i n'apareixen uns de nous.

Hi ha diverses maneres d'enfrontar-se a aquest problema, com veurem més endavant, però la més trivial és tenir un dispositiu de lectura de cada tipus disponible a tot moment. Està clar que quan deixi de funcionar, haurem de trobar un proveïdor que ens pugui aportar un de nou, però com que ja no és tecnologia actual, probablement el seu cost serà molt elevat. És per això que existeixen solucions millors.

Fins ara hem parlat del dispositiu lector del suport físic, però no és la única amenaça que es presenta en aquest àmbit. Hi ha un altre tipus de problema, molt més comú que no trobar lector compatible per al nostre dispositiu, i és que el propi suport físic hagi quedat deteriorat amb el temps, havent-se perdut tota la informació que en ell es contenia. És lo que anomenem "degradació dels suports".

Hi ha dos tipus de factors que influeixen en la degradació del suport. Els factors interns, referents a la pròpia construcció del suport físic, i els factors externs, referent a les condicions d'us i ambientals on s'emmagatzemen. Aquesta degradació es pot

mesurar en els anys que ens dura (de mitja) un determinat tipus de suport. Actualment, els dispositius òptics no solen durar molt més de 5 anys, mentre que, per exemple, encara avui podem llegir els jeroglífics escrits per els antics egipcis escrits tant al papir com a les parets de les tombes dels faraons. Està clar que hem guanyat en densitat d'informació, però hem perdut temps mitjà de vida.

Factors de degradació de suports	
Factors interns	**Factors externs**
Baixa qualitat dels materials	Falta de controls ambientals
Deficient procés de fabricació	Proximitat a camps magnètics
Causa temporal	Mal ús
	Mal estat dels equips de lectura

Taula 1. Factors de degradació de suports

Facilitat d'alteració de les dades

Per naturalesa pròpia, les dades digitals son molt fàcilment alterables, ja sigui com a conseqüència d'una errada als suports, o deterioració d'aquests, com el cas d'un us negligent o inclús un accés malintencionat. És important comptar amb control d'accessos a les dades i tenir-les replicades a vàries bandes per mitigar aquesta feblesa. A més un control de versions sempre ajuda a controlar els canvis produïts a un document i tornar enrere en cas de modificació errònia.

Elements a preservar de la informació digital

A diferència dels documents físics analògics, els documents digitals poden tenir diferents manifestacions, propietats o elements no estàndards. Això ens obliga a prendre decisions abans de preservar el document, com ara si volem preservar el document tal i com es va concebre en un principi, o només les dades essencials, sense més propietats. Per prendre aquesta decisió és important tenir en compte les necessitats d'informació dels usuaris actuals i futurs, i també el cost econòmic del manteniment.

Manifestació del recurs digital

A part de definir els elements que s'han de preservar, també hem de definir les formes en que el recurs es manifesta i quina d'aquestes ens interessa emmagatzemar. Hi ha 3 tipus de manifestacions del recurs:

- Binari: cadena de bits
- Lògic: dades codificades interpretables per un programari
- Conceptual: allò que l'usuari pot interpretar. Consisteix en un conjunt de fitxers digitals i metadades per reproduir una copia completa del objecte intel·lectual.

Elements interns del recurs

Les biblioteques estan comprovant com cada vegada els recursos digitals són més complexos i tenen més elements interns. Això augmenta la complexitat del pla de preservació en tant en quant s'han de prendre decisions damunt totes aquests elements.

Alguns d'aquests elements més complexos són:

- Funcionalitat
- Elements multimèdia
- Elements dinàmics
- Aplicacions interactives
- Hipervincles

Propietats essencials del recurs digital

Les propietats significatives o essencials d'un recurs van més enllà del contingut intel·lectual del recurs i fan referència a característiques com funcionalitat o apariència. Alhora de definir el nostre pla de preservació, haurem de definir què fer amb aquestes propietats. Tenim dues opcions:

- Mantenir el recurs tal i com es va dissenyar des d'un començament, amb el seu aspecte original, i amb les seves limitacions.
- Preservar el contingut i millorar la seva lectura incorporant noves utilitats disponibles en el moment de la preservació.

Autenticitat i integritat

Degut precisament a la facilitat d'alteració de les dades digitals, hem de tenir en compte mecanismes que ens permetin, no només conservar el document, si no també poder afirmar que és una còpia exacta del original i que és autèntic.

Podem definir autenticitat com la garantia del caràcter genuí i fidedigne del material digital, és a dir, que són lo que es diu d'ells que són.

Integritat és, per al contrari, que el estat dels objectes és complet i que no han sofert cap alteració no autoritzada.

Tot i que els conceptes poden estar ben definits, afirmar que un document digital és autèntic, no és tasca fàcil si no en disposam d'un original físic per comparar. I això serà en la gran majoria dels casos. Així doncs, la descripció proporcionada en el moment de la creació del fitxer, així com les seves metadades, són de vital importància per a assegurar el caràcter autèntic del document. En aquests casos, els usuaris que accedeixen a la informació han de tenir confiança en la institució que custodia els documents, en que ha fet correctament la seva feina de classificació i descripció, així com que ens assegura que el document no ha estat alterat de cap manera, i si ho ha estat, que estigui documentat correctament amb un control de canvis.

Anàlisi de riscos

Els anàlisis de riscos ens permeten analitzar i establir una solució per a tots els riscos que es pugin identificar en el nostre pla de preservació digital. La elaboració d'aquest pla és molt important ja que, com hem vist, un projecte de preservació està plagat de riscos i amenaces que posen en qüestió la seva viabilitat i els hem de donar una solució o com a mínim, conèixer que existeixen.

La presa de decisions i les solucions als riscos es basaran en:

- tipus de dany
- probabilitat de que es produeixi
- freqüència d'aparició
- repercussions
- opcions de prevenció

2.3. Aspectes no tècnics. Gestió de la preservació.

Tot i que els aspectes tècnics són els que a priori ens preocupen més quan parlem de la preservació d'un recurs digital, no hem d'oblidar que hi ha tota una sèrie d'aspectes relacionats que no són de tipus tècnic i que comprometen la viabilitat del nostre projecte de preservació. Aquests són:

- Aspectes organitzatius
- Aspectes econòmics
- Aspectes legals

Veurem que, en molts de casos, són aquests 3 aspectes els que compliquen la tasca de dur a terme un projecte de preservació digital, i no tant els aspectes tècnics esmentats al apartat anterior.

Aspectes organitzatius

Com ja hem vist abans, unes de les primeres passes a seguir en un pla de preservació és determinar què s'ha de preservar i qui ho ha de fer. Dins la organització que ha pres la decisió de dur a terme la preservació de recursos digitals hi haurà molts implicats, i és important que aquests actors sàpiguen quina és la seva funció dins el pla de la organització.

Durant la vida d'un document digital, aquest passa per vàries persones, com per exemple: el seu creador, les que ho han d'emprar, qui l'ha gestionat durant la seva vida útil, l'encarregat de la seva preservació, o el usuaris futurs, entre altres. Això no vol dir que tots aquests actors en siguin responsables directes de la preservació del document. S'ha d'establir correctament els rols de cada actor per tal de dur una correcta organització.

Si la preservació és una activitat que genera valor, probablement varis departaments de la organització se'n vulguin fer càrrec i haurà voluntat per dur a terme el projecte. Però si per al contrari és una obligació que no en genera valor per a la organització, o no es sap gestionar aquest valor, el projecte de preservació es veurà com una càrrega i no es donarà el suport adequat, fen que fracassi amb tota probabilitat.

Aspectes econòmics

Actualment és l'aspecte econòmic el que més dificultats presenta alhora de plantejar i dur a terme un projecte de preservació de documentació digital. A curt termini s'han de fer les inversions inicials en material i personal per a començar el projecte, però els

costos continuen al llarg termini per a mantenir tota la infraestructura existent i el creixent nombre de documents a incorporar al catàleg.

Un projecte de preservació sol necessitar d'una forta inversió inicial i no solen aportar grans beneficis econòmics a curt termini. Per això moltes institucions encara no donen suport a aquest tipus de projectes, o ho fan gràcies a subvencions de les administracions públiques com a part de projectes d'investigació. És important que cada vegada més, les organitzacions s'han d'adonar que els costos de preservació s'han de sumar a les despeses normals de la empresa, com un servei més, especialment una vegada finalitzada la subvenció pública.

S'estima que les despeses aproximades a un projecte de preservació digital es divideixen més o manco de la següent manera:

- 50% cost d'introducció de les dades
- 33% costs propis de la preservació
- 17% costs d'accés de les dades

Aquests costos com hem dit són aproximats i en dependran molt del tipus de projecte de preservació que es vulgui dur a terme. No obstant, es veu que, contràriament al que podríem pensar en un primer moment, no són els costos de preservació els que se'n duen la major part, si no els de introducció de dades. Això és per la gran quantitat de casos distints que ens podem trobar alhora de recepcionar un fitxer, i la quantitat de gent que involucra aquest procés, a vegades manual.

Les tasques pròpies de la preservació, encara que són una despesa constant, solen estar bastant automatizades i no tenen tant de personal involucrat, la seva despesa és bàsicament d'infraestructura i suports físics, els quals cada vegada tenen un preu per bit més reduït (llei de Kryder).

Per a minimitzar el costos de la fase inicial d'introducció de dades, s'han d'automatizar els processos al màxim possible, encara que mai s'assolirà un 100% d'automatització. També és convenient aplicar un filtratge als documents que volem preservar, la qual cosa reduirà convenientment el volum de dades a tractar i, a conseqüència, el seu cost.

Aspectes legals

Actualment la preservació digital es troba sense una regulació legal específica i s'ha d'ajustar a les lleis de propietat intel·lectual, molt restrictives en el que a còpies o canvis de format es refereix. Es necessita una regulació que llevi aquestes barreres a l'àmbit de la preservació, però de moment no existeix, obligant en molts de casos a

superar les barreres de forma il·legal, com ara llevant les proteccions o fent còpies no autoritzades del material digital.

Un altre problema legal al qual s'ha d'enfrontar un pla de preservació digital són els accessos regulats per a un determinat fitxer que necessiti una protecció especial per el seu contingut, com ara dades personals, mèdiques o econòmiques d'una població.

Dret d'autor

El dret d'autor és la protecció que s'ofereix al creador d'una obra reconeixent-li certs drets patrimonials i morals damunt ella. Mitjançant aquesta protecció es pretén fomentar la creativitat gràcies als estímuls de reconeixement i beneficis econòmics.

La propietat intel·lectual reuneix tant els drets d'autor com la propietat industrial. Així mateix, els drets d'autor es subdivideixen en dos tipus: drets morals i drets d'explotació.

El drets morals reconeixen, entre altres:

- Decidir si la obra ha de ser divulgada i en quina forma
- Determinar baix quin nom s'ha de publicar l'obra (real/pseudònim…)
- Exigir respecte a la integritat de la obra impedint la seva modificació

Per contra, els drets d'explotació, afecten al apartat econòmic, i descriu una sèrie d'activitats que no es poden dur a terme sense el consentiment del autor:

- Reproducció
- Distribució
- Comunicació pública
- Transformació

Aquests drets són atorgats en exclusiva al autor de la obra per un període de temps limitat. Actualment a la nostra legislació aquest temps és de 70 anys després de la mort del autor i a favor dels seus hereus. Una vegada finalitzat aquest termini, l'obra passa a ser de domini públic i es pot fer servir per qualsevol persona, amb finalitats comercials o no, sempre i quan es respectin els drets morals esmentats abans.

Així doncs, pel que fa a la preservació digital, els drets d'autor fan que necessitem del consentiment del autor o titulars dels drets implicats. Alguns actes de preservació digital que poden afectar als drets d'autor són:

- Còpia repetida per evitar la obsolescència del suport (refresc de suports). Acte de reproducció.
- Migració de la plataforma o del format. Acte de modificació de la obra.
- La migració pot afectar a la integritat de la obra, dret moral del autor.

2.4. Tècniques i estratègies de preservació.

Els sistemes de preservació digital fan feina amb un volum de dades molt important que requereix de personal especialitzat i que no pot escalar al mateix ritme que creixen les dades a introduir. Com varem veure, gran part dels costs dels sistemes de preservació se'n van a la fase d'introducció de dades, fent que es requereixi automatitzar al màxim totes les operacions possibles, i reduir la variabilitat tècnica dels documents a preservar, per tal de crear processos genèrics. Per això s'han de definir i emprar uns estàndards que augmentin la eficiència del sistema.

En aquest apartat veurem diverses solucions tècniques als riscos presentats al apartat anterior. Examinarem les principals opcions amb els seus avantatges i desavantatges, tenint en compte que cap d'elles resol totalment els problemes que ens han anat sorgint, per lo qual, en la majoria de casos, s'hauran d'aplicar en conjunt.

Refresc de suports

Anomenem refresc de suports o "refreshing" a la transferència de dades d'un suport a un altre de nou. Es fa per reduir el risc de pèrdua o alteració de dades degut al deteriorament físic del suport antic. Això és per que:

- Els suports tenen una vida útil limitada, encara que hagin sigut emmagatzemats en condicions òptimes.
- Els suports tornen obsolets degut a la inevitable evolució tecnològica, fent que sigui cada vegada més complicat trobar dispositius lectors de suports antics.
- Els suports poden sofrir alteracions físiques per causes externes o internes, com ara camps magnètics, humitat o canvis de temperatura.

La transferència de dades d'un suport antic a un de nou s'ha de fer sense canviar cap aspecte de les dades, ni el seu format. Només es refresca l'entorn, el contenidor, mai el contingut.

La freqüència de refresc de suports es farà d'acord amb les dades aportades pel fabricant i sempre tenint en compte les condicions ambientals d'emmagatzemament. En bones condicions ambientals, els suports poden arribar a durar molt més que en condicions dolentes. A més, no és bona idea intentar apurar al màxim les dades de durabilitat aportades pel fabricant ja que normalment, en la realitat aquestes solen ser menors.

El refresc de suports es pot fer entre el mateix tipus de suport o be, si ja s'ha produït un canvi generacional tecnològic dels tipus de suports, és recomanable fer-ho cap a

aquest nou tipus. Tot i així, s'ha d'estudiar el moment adequat per fer aquest canvi generacional. Hem de tenir en compte dues coses en canviar de tipus de format:

- Decidir si el format actual ja està obsolet o quedarà obsolet a curt termini. Això necessita d'un estudi de la evolució de la tecnologia actual, avaluant si està al final de la seva vida útil tecnològica o encara podem aguantar un altre cicle de refresc.
- Apostar per tecnologies consolidades. Si decidim canviar de tipus de format d'emmagatzemament per que considerem que el format actual ha quedat obsolet, hem de tenir ben clar que la nova tecnologia que adoptarem ha de tenir una gran acceptació com per a no haver de preocupar-nos de que quedi obsoleta molt prest o no trobem equips de lectura en els pròxims anys.

D'aquestes dues decisions en pot dependre que el pròxim refresc de dades surti més car o més econòmic, ja que els lectors de formats antics pugen de preu amb els anys, mentre que els lectors i suports surten més econòmics quan més acceptats estan.

El temps de refresc del suport és un altre aspecte clau a tenir en compte. S'ha de considerar que el canvi no és immediat i requereix un temps que no és despreciable, per la qual cosa, s'ha de planificar amb temps aquest canvi de suport físic.

Podem resumir les bones pràctiques en el refresc de suport en els següents punts:

- Utilitzar lectors i suports físics de qualitat
- Realitzar un bon us i manteniment dels equips
- Minimitzar riscos ambientals
- Establir polítiques de còpies de seguretat
- No apurar el temps de vida útil d'un suport.

Migració de formats

Amb la migració de formats volem convertir fitxers en formats antics a uns de nous.

Aquest canvi no és tan senzill i transparent com el del refresc de suports, en el que només canviava el contenidor i no el contingut. En aquest cas, canviar el format d'un fitxer per un altre compatible, sí hi ha una transformació del fitxer i de les seves dades, per això s'ha de comprovar al final del procés que el nou fitxer té el mateix contingut i propietats que el anterior.

La migració de formats no és una obligació com era el refresc de suports, en la que el fitxer podia quedar inservible si no es feia a temps. Moltes vegades, la migració de formats es fa de forma voluntària per obtenir qualcuns dels avantatges següents:

- Actualització a una versió més moderna del mateix format
- Passar d'un format propietari a un format lliure
- Emprar formats estandaritzats
- El format actual ha quedat obsolet i ja no hi ha software que el reconeix
- Per requeriment legal
- Unificació de formats distints a un únic format per facilitar la gestió

Tot i que es sol fer de forma voluntària, també hi ha casos en els que la migració és obligada, com ara quan ja no queda software actualitzat per fer servir amb el format, o el software funciona damunt un hardware obsolet. Per això és una estratègia que no s'ha d'oblidar, ja que no fer mai cap conversió de format pot derivar en la impossibilitat d'accedir a ell en un moment determinat en el futur pels motius esmentats. S'ha d'establir uns criteris per determinar quan un format està obsolet o està a punt de ser-ho i actuar en conseqüència.

Les migracions suposen un cost alt per a la organització, per això es fa necessari automatitzar al màxim el procés, per exemple, unificant els distints formats en un de sol, per agilitzar una futura migració. A més, s'han de tenir en compte els següents riscos:

- Pèrdua de funcionalitats o propietats significatives pròpies del format antic.
- Pèrdua de dades o introduir errades.
- Alt cost tècnic i humà.
- Alts temps de migració.

Davant als possibles canvis en la informació, funcionalitats o propietats significatives del document original, s'ha d'establir una estricta política de documentació de totes les migracions de format que es fan, de tal forma que si en trobem qualque dada a faltar, es pugui identificar o recuperar.

Degut al alt cost que pot suposar una migració de formats (nous sistemes, mà d'obra i molt de temps), i també degut als riscos que hem detallat, es recomana no fer moltes operacions d'aquest tipus. És per això que es recomana apostar per formats sòlids, estandaritzats i consolidats en la mesura de lo possible.

Hi ha dues qüestions finals que s'han d'esmentar en aquest apartat:

- El destí del arxiu original una vegada completada la migració de formats. Tenim dues opcions: guardar el fitxer original com a còpia del fitxer, o esborrar-ho per que ho hem migrat a un format millor. Si la migració s'ha fet de manera controlada i documentada, tindrem eines que ens permetran assegurar que el nou fitxer conserva totes les propietats del anterior i no és necessari guardar la

còpia original. Tot i que normalment la decisió presa és la de conservar el fitxer original, és una decisió insostenible a llarg termini en tant en quant suposa duplicar cada cert temps la informació del catàleg (i sense contar amb les noves incorporacions) la qual cosa fa que no sigui mantenible.

- S'ha d'avaluar la necessitat de fer migracions de formats de tot els fitxers encara que no hagin o no vagin a ser consultats a curt termini. Tenim la possibilitat de fer la migració de formats "en calent" quan un usuari demana un fitxer per consulta.

Varietats de migració de formats

Hi ha uns quants tipus de migració de formats, entre els quals en destaquem el següents:

- Migració sota demanda: es tracta de separar el codi binari de dades del seu entorn, enregistrant les metadades tècniques que especifiquen com s'han de migrar les dades cap a altres entorns. Així ens estalviem les migracions periòdiques i aquestes només es fan amb els documents que s'han de consultar.
- Migració a format estàndard: migració del document original a un format neutral o estandaritzat, per exemple PDF, ja que es considera que aquests formats tindran una vida més llarga que els formats propietaris no estandaritzats.
- Migració successiva: les dades s'actualitzen amb cada actualització del software, moment en el qual la conversió d'un format a un altre és més probable.

Emulació

Un dels principals motius pel qual es recomanava la migració de formats és la possibilitat que el fitxer ja no pugui ser llegit per cap software degut a que ja no funciona baix cap plataforma existent. Hi ha una altra solució per resoldre aquest problema: la emulació.

La emulació consisteix en crear un software que simuli el funcionament del hardware o software obsolet i així poder accedir al fitxer original amb aquest emulador, perfectament funcional amb les actuals plataformes.

L'avantatge d'aquesta tècnica és clara: no hi ha migració, no hi ha alteració del fitxer, ni pèrdua de temps en fer tot el procés, llevam tots els inconvenients de la migració. Molts autors creuen que la emulació és la solució més adient per als problemes de obsolescència de formats.

El major desavantatge que té la emulació és el cost econòmic i temporal de dissenyar i programar un software emulador per a cada un dels programes que llegeixen els formats obsolets.

Hi ha dos tipus d'emulació:

- Emulació de software: el software emulador, simula el funcionament d'un altre software que ja no és compatible amb els sistemes actuals.
- Emulació de hardware: el software que ha de llegir el document només funciona damunt un hardware obsolet. El emulador doncs, simula el hardware i el sistema antic.

Actualment s'està donant prioritat als emuladors hardware en front dels emuladors software. Això és degut a que el hardware antic és molt complicat d'obtenir, mentre que encara podem trobar còpies del software antic ben emmagatzemades a repositoris de software. Una vegada aconseguida la emulació del hardware, es pot instal·lar el software que només era compatible amb aquesta plataforma. A més, molts dels fitxers generats fa dècades, encara avui es poden obrir amb els programes moderns, que han anat arrastrant aquesta retro-compatibilitat amb formats ja obsolets.

Resum d'avantatges i inconvenients de la emulació:

Avantatges:

- Reprodueix el document original en la seva totalitat
- No és necessari un seguiment del format
- Permet visualitzar objectes complexos

Inconvenients:

- Poca experiència real
- Actualització dels emuladors amb els nous sistemes
- Programació complexa dels emuladors

Actualment, amb les eines de virtualització existents al mercat, tant de software lliure com propietari, lligat a les noves capacitats del hardware i sistemes moderns, moltes de les complicacions de la emulació desapareixen, però traslladen alguns dels seus problemes de complexitat a la programació del software de virtualització.

2.5. Eines per a la preservació

En aquest apartat veurem una selecció de eines que ajuden en la tasca de la preservació digital.

Comprovació de formats

Els sistemes de preservació han de conèixer el format en el qual estan codificats els seus fitxers i quins són els programes que en poder accedir. A la fase d'introducció dels fitxers al sistema de preservació, una de les primeres comprovacions que es fan, és la de comprovar el format del fitxer y enregistrar aquesta dada com a metadada.

Per tenir una referència de quin format és, si es conegut o no, i les seves característiques, és necessari comptar amb qualque Base de Dades de formats que sigui fiable.

Eines existents:

- Pronom és un registre de formats, de consulta pública, creat i actualitzat per els arxius nacionals de Regne Unit. El software que consulta les característiques d'un format determinat partir de les dades d'aquest registre, és DROID.
- La universitat de Harvard, als EEUU, hi ha un altre programa anomenat JHove que, a més d'identificar els formats dels fitxers, comprova que estiguin ben formats basant-se en les especificacions formals del estàndard del format.
- Unified Digital Format Registry, UDFR, de la universitat de California, que pretén ser una eina unificadora de totes les altres eines, per concentrar esforços només a una plataforma.

Comprovació d'integritat (*checksum*)

Aquesta eina ens permet comprovar que la informació continguda dins un fitxer digital no ha sofert cap alteració, ja sigui intencionada o no, des de la seva creació.

La comprovació d'integritat o *checksum* fa un càlcul d'una funció anomenada *hash* mitjançant la aplicació d'un algoritme matemàtic damunt el conjunt de bits del fitxer, obtenint al final del procés un valor únic. Les funcions de hash més emprades actualment són MD5, SHA-1 o SHA-256.

El càlcul de la funció de *hash* s'ha de fer en el mateix moment de la ingesta de dades, i s'ha d'emmagatzemar el valor resultant per a futures comprovacions d'integritat. Si en el transcurs del temps, ja sigui accidentalment o intencionadament, el fitxer canvia el seu contingut o estructura, els seus bytes també ho faran. Si es tornès a calcular la funció de *hash* damunt aquest nou fitxer, ens donaria un valor diferent al obtingut en

primera instància a la ingesta del fitxer, indicant que aquest fitxer no és igual al anterior. Aquest tipus de comprovacions han d'estar integrades dins qualsevol programa de gestió d'un magatzem digital, ja que és una eina ràpida i que ens aporta seguretat damunt la integritat del fitxer.

Transferència de fitxers

Moltes entitats que comencen un projecte de preservació digital no tenen la suficient capacitat tècnica, organitzativa ni econòmica per executar un pla de preservació a llarg termini per elles mateixes. Això fa que hagi d'externalitzar aquest servei en altres empreses o institucions. Altres vegades és un impositiu legal, o simplement col·laboració entre entitats o redundància de fitxers a diferents magatzems digitals.

Per a assolir això, la primera passa, és transferir els seus fitxers d'un magatzem origen cap al magatzem de dades destí que les preservarà. Per garantir que la transferència es fa correctament i amb garanties, hi ha una sèrie d'estàndards de transferència de fitxers. Aquests estàndards defineixen el format dels fitxers a transferir, proporcionant un entorn normalitzat, i especifiquen els protocols de transmissió de dades.

Alguns d'aquests protocols són:

- JATS: és un tipus d'esquema XML creat per la National Library of Medicine dels Estats Units per a codificar la estructura de les revistes científiques i fer la transmissió entre editorials i clients.
- Bag-It: és un programari desenvolupat per la Libray of Congress dels Estats Units i la California Digital Library per transmetre grans volums de dades a centres de preservació digital.

Aquests estàndards d'empaquetat faciliten la transmisió i la ingesta automatitzada de les dades en el sistema receptor, facilitant i abaratint la tasca de introducció de dades.

Metadades

Junt a les dades que s'han de preservar, s'han de reunir tot un conjunt de dades descriptives de la obra en qüestió. Aquestes dades seran fonamentals en la catalogació i posterior cerca al magatzem de dades. Són les anomenades "metadades". La seva importància és vital als sistemes de preservació, tant analògics com digitals, ja que aquests requereixen d'un sistema de descripció uniforme, consistent i de llarga durada per a ser útils quan enregistren un gran volum de dades.

Hi ha diverses classificacions de les metadades depenent del autor que es consulti, però una classificació més o manco estesa és la següent:

- Descriptius: orientats a la descripció intel·lectual del contingut
- Tècnics: orientats a la descripció tècnica del contingut
- Propietat intel·lectual: orientats a la definició dels drets i propietaris del fitxer

Metadades descriptives

Són les metadades que descriuen el contingut del fitxer preservat, aportant referències i informació per a una correcta identificació i posterior recuperació. Sense aquestes metadades, no sabríem què estam preservant, i molt probablement ningú torni a accedir a aquest contingut per no conèixer el seu contingut.

Una definició més precisa, aportada per el NISO diu que les metadades són "*informació estructurada que descriu, explica, localitza o facilita la recuperació, ús o gestió de un recurs d'informació*".

L'esquema de metadades més conegut i implementat en les biblioteques és el Dublin Core. Aquesta iniciativa va sorgir d'una reunió d'experts en la matèria celebrada en 1995 a la ciutat de Dublin, Ohio. En 1998 varen publicar el Dublin Core Metadata Element Set que consisteix en 15 elements considerats de caràcter bàsic per a tot tipus de material digital. Al 2001 es va adoptar con a norma ANSI, i al 2003 com a norma ISO baix el nom ISO 15836:2003. Aquests 15 elements que descriu són els següents:

1. Títol
2. Autor/s
3. Claus (o descriptors)
4. Descripció
5. Editor
6. Altres col·laboradors
7. Data
8. Tipus de recurs
9. Format
10. Identificador del recurs
11. Font
12. Idioma
13. Relació
14. Cobertura
15. Drets de propietat intel·lectual

A partir d'aquesta llista bàsica, cada sector pot adaptar els elements als seus propis usos especialitzats i completar-los amb altres dades.

La norma METS (Metadata Encoding and Transmission Standard) és un esquema XML per a codificar les metadades dels objectes digitals. Aquesta norma explica com codificar les metadades descriptives, de propietat intel·lectual, tècnics (com ara el hash, la resolució, la mida en bytes…) i estructurals (quins fitxers representen l'objecte i la seva relació).

La norma METS és extensible i permet que es puguin codificar les metadades segons els estàndards del Dublin Core, MARCXML o PREMIS.

Metadades de preservació

Les metadades de preservació permeten que els programes de preservació puguin comprendre com representar els materials digitals quan el seu accés sigui necessari. Aquest tipus de metadades són imprescindibles en qualsevol programa de preservació. Són metadades descriptives, estructurals i administratives que documenten la preservació a llarg termini dels recursos digitals.

Les metadades de preservació permet que un recurs digital s'auto-documenti durant qualsevol fase del seu cicle de vida. Les dades que constitueixen aquesta auto-documentació són:

- Característiques i entorns tècnics
- Procedència del recurs
- Activitats de preservació realitzades per a preservar aquest recurs
- Validació de autenticitat
- Drets d'autor

PREMIS és un esquema XML que implementa un catàleg de metadades de preservació. El seu objectiu, com les metadades de preservació, és de documentar, d'una forma estandaritzada, aquells aspectes que puguin afectar a la preservació d'un recurs digital. Alguns exemples d'esdeveniments que es documenten amb PREMIS són:

- Migracions de formats
- Canvis a la política de *hash*
- Controls de seguretat
- Controls d'integritat
- Canvis als drets de propietat intel·lectual

La importància de PREMIS està dins el seu Diccionari de dades, que defineix diferents entitats, com ara:

- Entitat intel·lectual: continguts tractats com una unitat

- Objecte: unitat discreta d'informació en format digital
- Esdeveniment: registren la procedència digital i contribueixen a demostrar la autenticitat del objecte
- Agent: persona, software o institució associada als esdeveniments durant la vida d'un objecte.
- Drets: declaració de drets o permisos d'un objecte o agent.

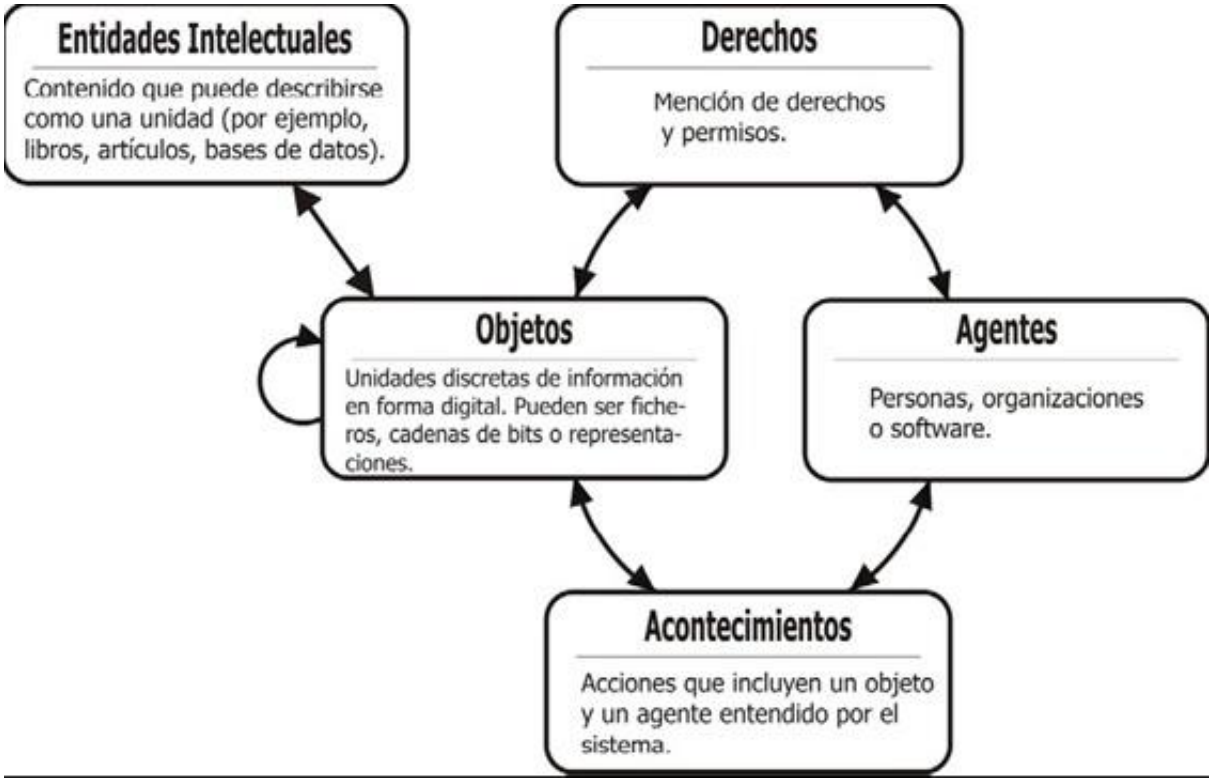


Diagrama 2. Model de dades de PREMIS on es defineixen les 5 entitats i les seves relacions.

Emmagatzemament de metadades

Les metadades de preservació poden quedar encapsulats dins els mateix recurs digital o be, emmagatzemar-se en un fitxer separat. Hi ha arguments a favor i en contra de cadascuna de les dues opcions.

Incrustació de les metadades

- A favor: encapsular les metadades dins el propi recurs digital fa que aquest s'auto identifiqui. Quan es recuperi el recurs en el futur, es disposarà de la informació necessària per a identificar-lo i interpretar-lo.
- En contra: intervenció repetitiva en el recurs en les operacions d'actualització.

Metadades independents

- A favor: facilitar la localització, manipulació i ús del recurs si aquest no està disponible a una xarxa pública.
- En contra: els fitxers que contenen les metadades també són recursos digitals que s'han de preservar juntament amb l'original.

2.6. Model OAIS

El model OAIS és un model teòric que indica quines funcions han de suportar els sistemes de preservació digital, sense tenir en compte el tipus de dades ni l'activitat u organització al quals es refereixen. No és un software, un hardware ni un format. OAIS és un model que descriu 6 grans blocs de processos dins un arxiu de preservació digital.

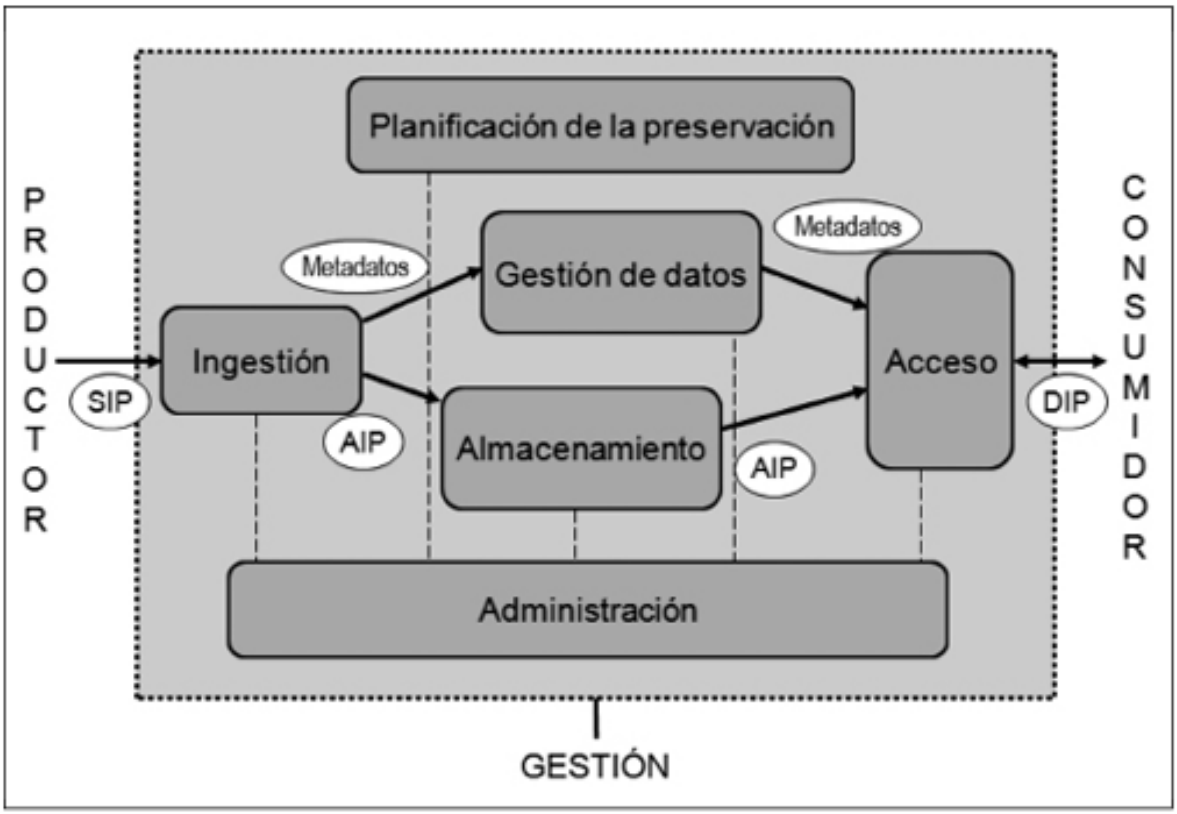


Diagrama 3. Model OAIS

1. Ingesta: el primer bloc s'encarrega detractar els fitxers que arriben provinents dels productors (oficines, sistemes informàtics…) i se'ls aplica una sèrie de controls previs com ara Control de procedència, antivirus o control de formats. Una vegada fetes aquestes comprovacions el sistema extreu les metadades de tipus tècnic i calcula la clau de *hash* per a comprovació d'integritat. Altres activitats realitzades en aquest punt són:
    a. Assignar identificador únic al recurs
    b. Seleccionar la estratègia tecnològica basada en les propietats essencials del recurs
    c. Convertir el recurs i les seves metadades en una cadena de bits
    d. Elaborar la Representació de la Informació (RI): dades tècniques necessàries per a convertir l'objecte digital a altres formats.

    e. Elaborar la Descripció de la Informació de Preservació (PDI): dades descriptives del objecte.
2. Emmagatzemament: procés encarregat del emmagatzemament del recurs digital als suports físics, normalment arrays de discos amb protocols de còpies de seguretat i redundància de dades. Altres activitats d'aquest bloc són:
    a. Passar els recursos des de la fase anterior (ingesta) cap a la d'emmagatzematge permanent
    b. Gestionar el sistema d'emmagatzematge
    c. Comprovacions d'integritat de les dades i refresc de suports
    d. Subministrar la informació necessària per a l'accés i difusió del recurs
3. Gestió de les dades: en aquest procés s'emmagatzemen les metadades, tant les pròpies del fitxer com les calculades a la fase 1 d'ingesta de dades. L'objectiu és disposar de tota la informació centralitzada i normalitzada. Aquest procés anirà enregistrant totes les incidències que afectin al recurs digital en qüestió, com ara migracions de formats o alteracions d'integritat. Altres activitats d'aquest apartat són:
    a. Control del accés
    b. Seguiment de les sol·licituds dels usuaris
    c. Mantenir els perfils dels usuaris
    d. Mantenir les dades relacionades amb la seguretat dels recursos (com ara noms d'usuari, passwords...)
    e. Estadístiques del funcionament del sistema
    f. Dades contables
4. Accés: aquest procés habilita mecanismes d'accés als recurs preservats per part dels usuaris. A partir de la informació demanda per l'usuari, farà una cerca a les metadades enregistrades a la Gestió de dades i recuperarà els fitxers corresponents del magatzem de dades. Aquest procés també s'encarrega de comprovar les credencials del usuari i determinar si té permís per accedir al recurs digital sol·licitat. Aquest bloc ha de proporcionar una interfície per a mostrar les dades i fer una migració de formats "en calent" si escau. A més, també s'encarrega de:
    a. Comprovar la integritat del recurs demanat per l'usuari
    b. Validar el nivell d'autorització per a accedir a un determinat recurs
5. Preservació: aquest procés és on es decideix quines polítiques de preservació es duran a termini. Els responsables del sistema de preservació vigilaran tots els perills citats a capítols anteriors, com ara la caducitat d'un format determinat o la disponibilitat de noves eines per accedir al recurs. També s'encarrega de:
    a. Fer seguiment de les novetats tecnològiques
    b. Desenvolupar estratègies per assegurar l'accés continuat
    c. Desenvolupar plans de transferència i migració de dades
6. Administració: suports tècnics comuns a tot els altres processos.

a. Negociar acords amb els proveïdors dels recursos
b. Revisió de procediments vigents
c. Actualitzar les polítiques de del dipòsit
d. Suport als usuaris
e. Mantenir contactes amb altres repositoris i institucions

Transferència de les dades entre els blocs de processos

OAIS també determina com s'han de moure les dades entre cadascun dels blocs de processos abans mencionats. Això és important per que sovint aquests blocs estan constituïts per sistemes informàtics diferenciats o inclús a ubicacions geogràfiques distintes.

- SIP (Submission Information Package): paquet de dades que inclou els fitxers de dades que s'envien al sistema de preservació, amb les metadades d'integritat i autenticitat.
- AIP (Archival Information Package): paquet similar al anterior però s'aplica a la comunicació entre els processos de Ingesta i Emmagatzemament.
- DIP (Dissemination Information Package): fitxers de dades que reben els usuaris com a resposta d'una petició al sistema. Inclou metadades de integritat i autenticitat a més de metadades de la història del recurs.

3. Cas d'estudi d'una biblioteca universitària

Introducció

El present cas d'estudi tractarà de definir un projecte de preservació digital per al fons bibliogràfic d'una biblioteca universitària.

En primer lloc s'haurà de definir el pla de preservació, estudiant les possibilitats a l'abast de la nostra organització, definint el tipus de preservació que es vol implantar, quins materials seran objecte de preservació i les persones implicades.

Una vegada definit el pla, s'haurà d'implantar, establint uns criteris de nivell de servei, i fixant les tasques de manteniment. També s'hauran d'estudiar possibles modificacions al pla per adaptar-lo a les futures necessitats de la organització, així com a un ambient econòmic cambiant.

En aquest cas pràctic, s'estudiaran diferents alternatives per a alguns conceptes i el seu impacte en la organització, ja sigui a nivell funcional, organitzatiu i econòmic. També podrem basar les nostres decisions en el que hagin fet altres institucions, estudiant la seva solució i mirant d'aplicar-la al nostre projecte en cas que sigui possible.

S'ha de tenir en compte que cada pla de preservació, encara que segueix un model comú, o unes tècniques més o manco similars, és únic, especialment entre institucions de diversos sectors. No seran les mateixes necessitats les d'una administració pública que les d'una editorial o les d'una productora audiovisual.

Basarem la elaboració del nostre pla de preservació al voltant de les següents qüestions:

- Definir i implementar una política de preservació
- Definir els procediments de gestió
- Definir una infraestructura per al magatzem de dades
- Identificar i seleccionar els recursos a preservar
- Aspectes econòmics i financiació
- Definir el començament del projecte i la seva mantenibilitat

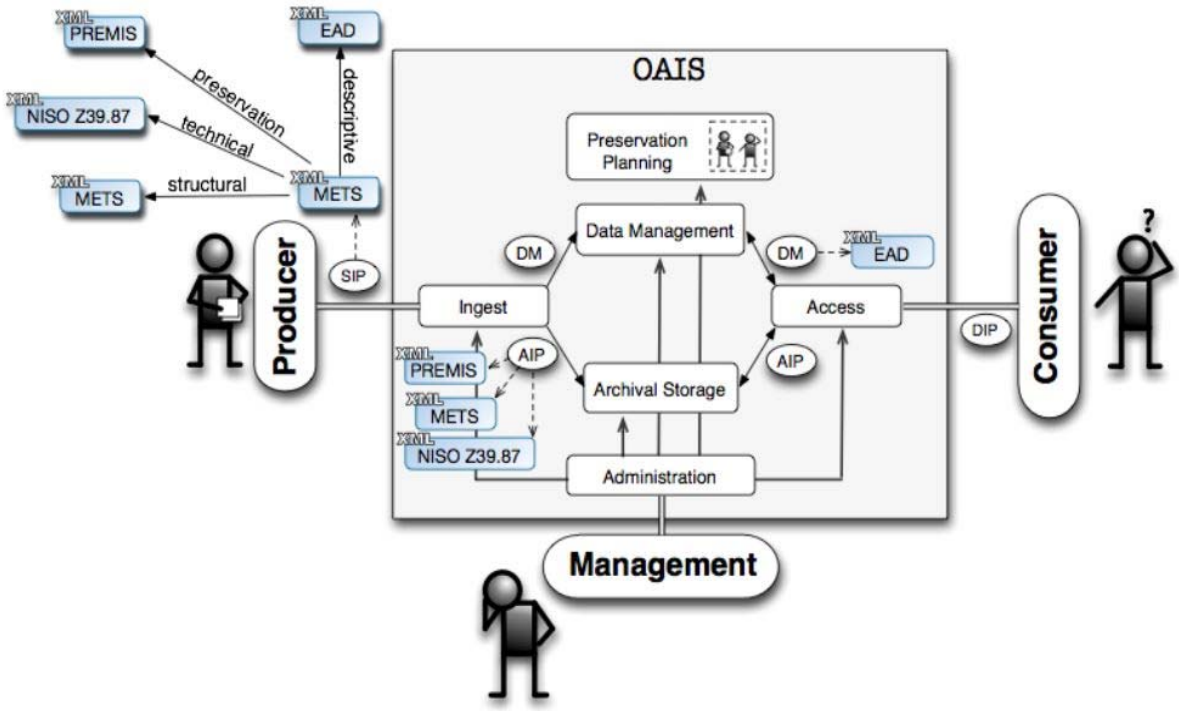


Diagrama 4. Model d'informació OAIS expandit

Com a base del projecte agafarem el model de referència OAIS que és també una norma estandarditzada per ISO (ISO 14721). Així estructurarem aquesta part seguint cadascun dels 6 apartats que té el model:

I. Planificació de la preservació
II. Administració i gestió del dipòsit digital
III. Ingesta
IV. Magatzem de dades
V. Gestió de les dades
VI. Accés

3.1. Planificació de la preservació

En aquest apartat definirem el nostre pla de preservació, el seu abast, els recursos que hi destinarem o la elecció dels fitxers a incloure en el nostre sistema.

Recursos a preservar

Una de les primeres tasques a dur a terme en el nostre sistema és la de definir quins seran els recursos que s'han d'incloure dins el pla de preservació. Aquests vindran definits per una sèrie de paràmetres com ara:

- Obligatorietat de preservació per imperatiu legal
- Valor de la obra
- Política institucional
- Permisos d'accés al material
- Disponibilitat del recurs

En el nostre cas, la Universitat a la qual pertany la nostra biblioteca vol preservar els següents tipus de documents:

- Obres publicades per la editorial universitària
- Treballs de final de màster i projectes de final de carrera
- Fons documental adquirit per la universitat
- Butlletins i publicacions oficials dels organismes rectors de la universitat
- Publicacions web de la universitat i dels professors com ara blogs, tweets...
- Apunts penjats al sistema de e-learning de la universitat
- Documents audiovisuals creats per el laboratori multimèdia
- Software desenvolupat per la Universitat
- Software específic necessari per fer a la docència
- Bases de dades
- Dades científiques generades per els equips d'investigació
- Dades acadèmiques recollides per la Universitat (com ara llistes d'alumnes o qualificacions)

Aquesta llista no és exhaustiva i hi poden aparèixer nous tipus de recursos a preservar en el futur. Per això s'ha d'estar vigilant periòdicament la producció documental de la biblioteca i la universitat, així com les noves adquisicions de material.

Com es veu, els tipus de recursos a preservar poden ser molt variats, des de documents text, a pàgines web, vídeos o inclús software. Per tal d'estandarditzar les entrades, s'han de definir quins són els tipus de fitxers i les extensions que el sistema

admet per a cada un d'ells i si són preferits, acceptats o problemàtics. S'han definit les següents:

Tipus de recurs	Format preferit	Format acceptat	Format problemàtic
Tipus text	.txt .odt .rtf	.doc .docx	
Imatges Bitmap	.tiff .png .jpg	.gif	.bmp
Imatges vectorials	.svg .odg	.dwg .dxf	.ai .swf
Àudio	.mp3 .ogg .midi	.wav	.cda
Vídeo	.mpg .ogv	.mp4 .mkv .avi	.mov .wmv
Fulls de càlcul	.ods .csv	.xls .xlsx	
Bases de dades	.sql .csv	.mdb .accdb	
Obres digitalitzades	tipus text	.pdf	tipus imatge

Taula 2. Classificació de formats segons la seva preferència al nostre sistema.

En general, es prefereixen els formats oberts als tancats, per que hi sol haver més programari disponible per a la seva utilització a més que aquest programari sol ser gratuït (encara que no sempre és així). Si no es pot aconseguir el recurs digital en format lliure, es prefereixen els formats tancats però àmpliament emprats (estàndards *de facto*), com ara els de Microsoft o Adobe. Els formats tancats i poc estesos, amb poc suport, són els formats que ens poden portar més problemes i per això és convenient detectar-los i, en la mesura de lo possible, convertir-los a un dels altres formats. Per exemple, si tenim un vídeo enregistrat amb Quicktime (.mov), abans de ficar-lo dins el sistema de preservació, es preferirà fer la conversió a un format de més fàcil tractament. Això es tractarà al apartat de la fase d'ingesta.

En el cas de projectes software, si han sigut desenvolupats per la pròpia Universitat, serà desitjable no només enregistrar el producte final (programari executable) si no també, en la mesura que es pugui, el codi font del mateix, el qual normalment no necessita un tractament especial (solen ser fitxers de text on hi ha el codi o la configuració del projecte) i no ocupa gaire més.

En quant a les obre digitalitzades, es seguirà el mateix criteri de selecció que el de les obres natives digitals, però tenint en compte que serem nosaltres, a la fase d'ingesta, qui triarem com es fa aquesta digitalització, així que triarem el format més adient per

al nostre pla de preservació, en aquest cas, tipus text o PDF amb cerca OCR. En parlarem més endavant, a la fase d'ingesta.

Addicionalment, i com veurem més endavant, per tal d'optimitzar els recursos entre organitzacions, es poden consultar a altres entitats el seu compromís de preservació amb material que nosaltres anem a preservar. Si trobem que altres biblioteques o institucions ja inclouen determinat material, podríem establir acords de col·laboració amb ells per tal de estalviar-nos aquestes obres i accedir als seus sistemes per a consultar aquestes obres.

Planificació temporal del pla de preservació

Un dels aspectes més importants del nostre pla és decidir quan començar a implementar-ho i a partir de quin moment s'han d'incloure les obres dins el nostre sistema. També s'ha de preveure el creixement que tindrà el nostre fons documental amb el temps.

En el nostre cas d'estudi, la biblioteca universitària no té cap projecte de preservació en marxa, així que partim de zero. Per tant, s'ha decidit posar en marxa el projecte quan més aviat possible. Això s'ha decidit per que la organització és conscient dels riscos de no fer res amb la preservació de les dades, com ara: pèrdua de la informació, impossibilitat d'accés futur… etc.

Actualment la única actuació que té en producció el sistema documental de la biblioteca són còpies de seguretat del seu fons digital. Aquestes còpies es fan setmanalment a altres suports i s'emmagatzemen a un edifici diferent al dels originals.

Al nostre pla de preservació definirem que el abast temporal del projecte es indeterminat, això és, no té fi definit, per que, volem que els documents i recursos que hi posem puguin estar disponibles indefinidament. No definirem una data de caducitat per als recursos, tot i a que a altres plans de preservació es podrien definir per a determinat recursos. En el nostre cas, com tot el material és acadèmic, encara que amb el temps el contingut quedi obsolet, volem tenir-ho com a referència del estat actual de les investigacions i estudis. Això tindrà, com veurem més endavant, un impacte dins la gestió econòmica i tècnica del pla de preservació.

En aquest apartat també definirem els tipus d'actuacions que durem a terme al nostre pla de preservació tal i com varem veure a la part teòrica.

Actuació reactiva Vs Actuació proactiva. Està clar que en el nostre cas el que tenim ara mateix és una actuació reactiva: ens hem topat amb el problema de la preservació de les dades i hi ha gran quantitat de fons documental fora de cap tipus de preservació.

Durant les primeres fases del projecte, la nostra actuació serà purament reactiva. Però una vegada incorporat tot el fons actual al projecte de preservació, s'ha d'actuar proactivament, això és, cada vegada que entri un nou document a la nostra biblioteca, sempre que sigui d'uns dels tipus a tractar, s'incorporarà al sistema lo més aviat possible, i no deixar-los a una cua de pendents.

Actuació esporàdica Vs Actuació continuada. Si volem que el nostre projecte de preservació sigui seriós, les accions han de ser continuades. Tal i com hem dit abans, no és gaire bona idea disposar d'una cua de recursos pendents d'introduir al sistema per que això demostra que no s'està planificant bé la ingesta de documents o no s'ha calibrat correctament la mida dels recursos a incloure dins el sistema.

A la planificació temporal del nostre projecte s'han d'avaluar moltes variables. En detallarem les més importants:

- Definició i abast del projecte. Aproximadament 1 mes.
- Selecció i contractació de personal especialitzat: 3 setmanes.
- Ingesta de documents
    - digitalització: de 8 a 10 mesos.
    - migració de formats i integració dels actuals documents existents: 1 mes
- Posar en marxa el sistema d'emmagatzematge físic: 1 mes
- Desenvolupament del sistema d'extracció i emmagatzematge de metadades: 2 setmanes.
- Desenvolupament del sistema de accés, visor i emuladors: 2 mesos.
- Desenvolupament del sistema de seguretat, usuaris i rols: 2 setmanes.

Id		Nombre de tarea	Duración	Comienzo	Fin
1		Definició del projecte	1 ms	mié 08/06/16	mar 05/07/16
2		Aprovació del projecte	0 días	mié 06/07/16	mié 06/07/16
3		Contractació nou personal	3 sem.	vie 08/07/16	jue 28/07/16
4		**Desenvolupament**	**40 días**	**mié 20/07/16**	**mar 13/09/16**
5		Sistema d'emmagatzematge	1 ms	mié 20/07/16	mar 16/08/16
6		Sistema d'extracció de metada	2 sem.	mié 17/08/16	mar 30/08/16
7		Sistema visor	2 mss	mié 20/07/16	mar 13/09/16
8		Sistema de seguretat	2 sem.	mié 17/08/16	mar 30/08/16
9		Tests	1 sem	jue 15/09/16	mié 21/09/16
10		Posada en marxa	0 días	mié 21/09/16	mié 21/09/16
11		**Ingesta**	**180 días**	**lun 26/09/16**	**vie 02/06/17**
12		Digitalització	9 mss	lun 26/09/16	vie 02/06/17
13		Migració de formats	1 ms	lun 26/09/16	vie 21/10/16

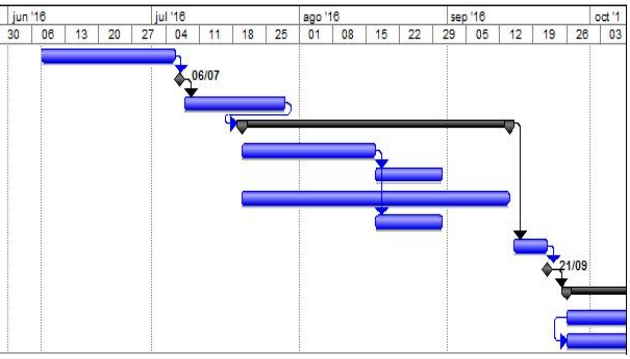


Diagrama 5. Gantt per a la planificació temporal del projecte de preservació. La tasca de digitalització surt tallada per la seva longitud.

Persones i institucions implicades

Totes les persones implicades en el actual procés de preservació digital, per primitiu que sigui (recordem que només es feien còpies de seguretat) i les persones responsables de la custodia dels elements analògics actuals a la biblioteca, s'han d'implicar dins el nou pla de preservació digital. Hauran d'adquirir noves competències i en alguns casos haurem de contractar personal amb coneixements més específics per a la tasca de la preservació digital. Les institucions com ara les editorials o la pròpia universitat també s'han d'incloure com a implicats procés.

Persones i institucions implicades:

- Cap de biblioteca: és el que coneix l'actual sistema i el responsable del àrea. La seva col·laboració és essencial. En el nostre cas, a més, va ser idea seva la de implantar un vertader sistema de preservació digital, així que la seva implicació està assegurada. És ell qui elaborarà el pla final de preservació junt amb els tècnics de la Universitat i el personal jurídic.
- Personal tècnic: assessoraran al cap de biblioteca en la elaboració del pla de preservació, aportant el coneixement tècnic per al magatzem de dades. Una vegada comenci la fase d'implantació, es dedicaran a la implementació de les solucions triades. Quan el sistema estigui operatiu i estable, hauran de fer tasques de manteniment i millores.
- Bibliotecaris: com a persones de contacte amb els usuaris, podran donar suport i ajuda a aquests amb el sistema de preservació mitjançant els accessos web que es defineixen.
- Gerència / Rectoria: com a responsables econòmics de la universitat, la seva implicació en el projecte és clau. El sistema de preservació té un costs importants de implantació i tal i com l'hem definit, necessitarà de partides pressupostàries anuals de sostenibilitat per un període de temps indefinit.
- Editorials: tant la editorial universitària com les altres editorials a les quals adquirim fons bibliogràfic, han de poder aportar les seves obres en format digital en el millor format per a la preservació i establir acords legals pels quals el nostre sistema pugui custodiar les seves obres.
- Personal jurídic: ajuden en el assessorament legal del pla de preservació i revisen aquest si hi ha qualque canvi normatiu que pugui afectar al llarg del temps.
- Altres universitats: podem establir acords de col·laboració amb altres entitats universitàries o institucions públiques que també hagin implantat un sistema de preservació digital per demanar consell o assessorament, i també per informar-nos del fons que ells estan preservant per tal de no duplicar esforços. Com hem dit abans, si trobem que altres entitats preserven els mateixos recursos que nosaltres, podem trobar d'establir un acord de col·laboració per

tal d'accedir al seu fons i estalviar-nos la preservació dels mateixos. Això sí, s'ha de fer un seguiment periòdic d'aquests recursos per tal de que, si deixen d'estar disponibles al altre fons, poder iniciar la seva preservació al nostre sistema.

Anàlisi de riscos

En aquest apartat detallarem els riscos als quals s'enfronta el nostre sistema de preservació i els intentarem donar una solució, si es que en tenen. S'ha de dir que riscos n'hi ha molts, alguns molt poc probables, i altres de desconeguts, per això a la següent taula mostrarem aquells que tenen més probabilitat d'aparèixer segons el nostre entorn.

A la següent taula en mostrem alguns d'aquests riscos:

Risc	Probabilitat	Repercussions	Actuacions
Errada al suport físic. Degradació.	Elevada	Inutilitza el suport i perdem les dades que hi conté.	Replicació; suports de millor qualitat.
Obsolescència del suport	Mitjana	Tot suport queda obsolet tard o d'hora. Perdem l'accés a les dades.	Refresc de suports quan detectem que la vida útil de la tecnologia està acabant.
Errada elèctrica	Mitjana	Perdem l'accés a les dades temporalment	Inclusió de sistemes d'alimentació ininterrompuda (SAI). Sistema replicat.
Inundació o desastre natural	Baixa	Inutilitza el suport i les màquines. Alt cost de reparació.	Replicació a llocs geogràfics diferents.
Accés malintencionat	Baixa	Pot inhabilitar gran quantitat de dades.	Controls de seguretat, control d'accés, verificació hash, control de versions.
Mal funcionament del software	Baixa	Pot inhabilitar gran quantitat de dades.	Controls de qualitat del software, tests, verificació hash, control de versions.

Canvis legals	Baixa	Obligació legal d'incloure recursos o de llevar-los del nostre repositori	Apel·lació si es possible.
Retalls econòmics	Mitjana	Posa en perill la continuïtat del projecte	Reducció de recursos a preservar.

Taula 3. Anàlisi de riscos

Anàlisi de creixement

Un sistema de preservació digital sempre va rebent noves dades i nous recursos per a ser incorporats, mentre que manté els recursos antics. Això fa que el sistema no deixi de créixer mai, ja que, en el nostre cas, no contemplem la possibilitat de que els recursos caduquin i s'hagin de llevar del sistema de preservació.

En una fase inicial, on ficarem dins el sistema tots els recursos digitals que trobem i que compleixin els criteris establits per la organització, el creixement serà molt ràpid i lineal. La velocitat de creixement vindrà donada per la nostra habilitat a la hora d'introduir els recursos dins el sistema. Si només comptem amb medis humans, aquest procés es pot allargar bastant. Si per el contrari, fem us de qualque eina d'automatització de la tasca, el temps es pot reduir considerablement. És important trobar o dissenyar una eina que ens ajudi a la introducció inicial de les dades.

Així mateix, a la fase inicial també hi haurà molta feina de digitalització de tot el fons analògic que volem preservar. Aquesta tasca, encara que es pot automatitzat, sol ser més lenta i requerir de personal altament especialitzat, especialment en el cas d'obres úniques o molt antigues.

Actualment el fons de la biblioteca consta d'un total de 75.000 obres, aproximadament, incloent llibres, treballs, revistes, DVD's, CD's i arxius digitals.

L'actual ritme de creixement de la biblioteca estava entorn a les 2.000 obres anuals, degut a les noves publicacions universitàries i a les adquisicions que es feia amb el pressupost assignat.

En el cas que s'ha d'incloure material digital addicional, com ara software, imatges, projectes de fi de màster… es calcula que el creixement podria rondar les 20.000 obres anuals. I es preveu que el creixement sigui major cada any degut a la major quantitat de publicacions digitals que es fan any a any.

Aquest ritme de creixement en obres, i en bytes, condicionarà l'infraestructura tecnològica i el pressupost assignat al pla. És probable que, depenent del pressupost

assignat, s'hagin de reduir les obres a incloure dins el pla de preservació o un pla de caducitat de certes obres que perden interès amb el pas dels anys.

Solució comercial Vs desenvolupament propi

En aquest punt, i amb part del pla ja desenvolupat, s'ha de tractar una qüestió de vital importància per al sistema de preservació: emprar un sistema comercial que cobreixi la major part de les nostres necessitats, o desenvolupar un software específic per al nostre cas en concret.

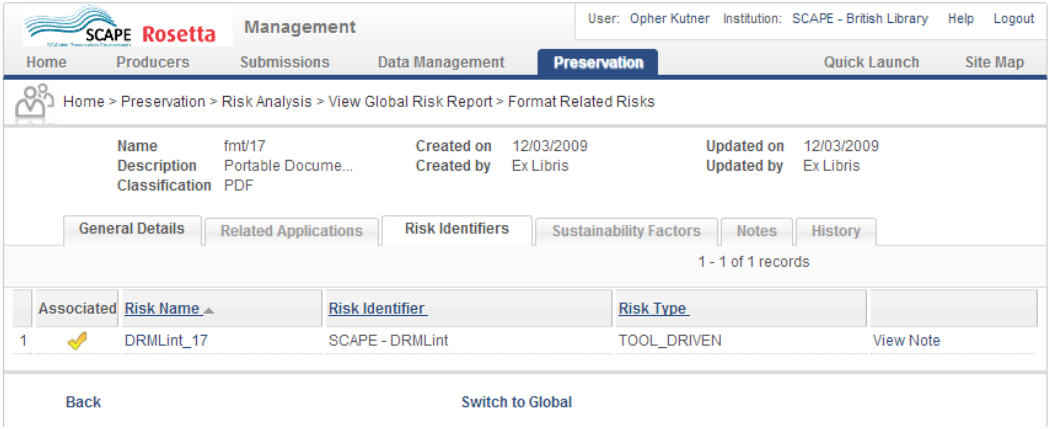
La primera solució té els avantatges que ja està desenvolupat i es més ràpid de posar en marxa el projecte, a més que es una solució ja provada per altres institucions, amb menor probabilitat de trobar errades en un principi.

La segona solució, probablement més costosa en temps i diners en un primer moment, ens permet desenvolupar una solució perfectament a mida a les nostres necessitats.

Veurem quines solucions comercials hi ha al mercat, el seu preu i les seves prestacions.

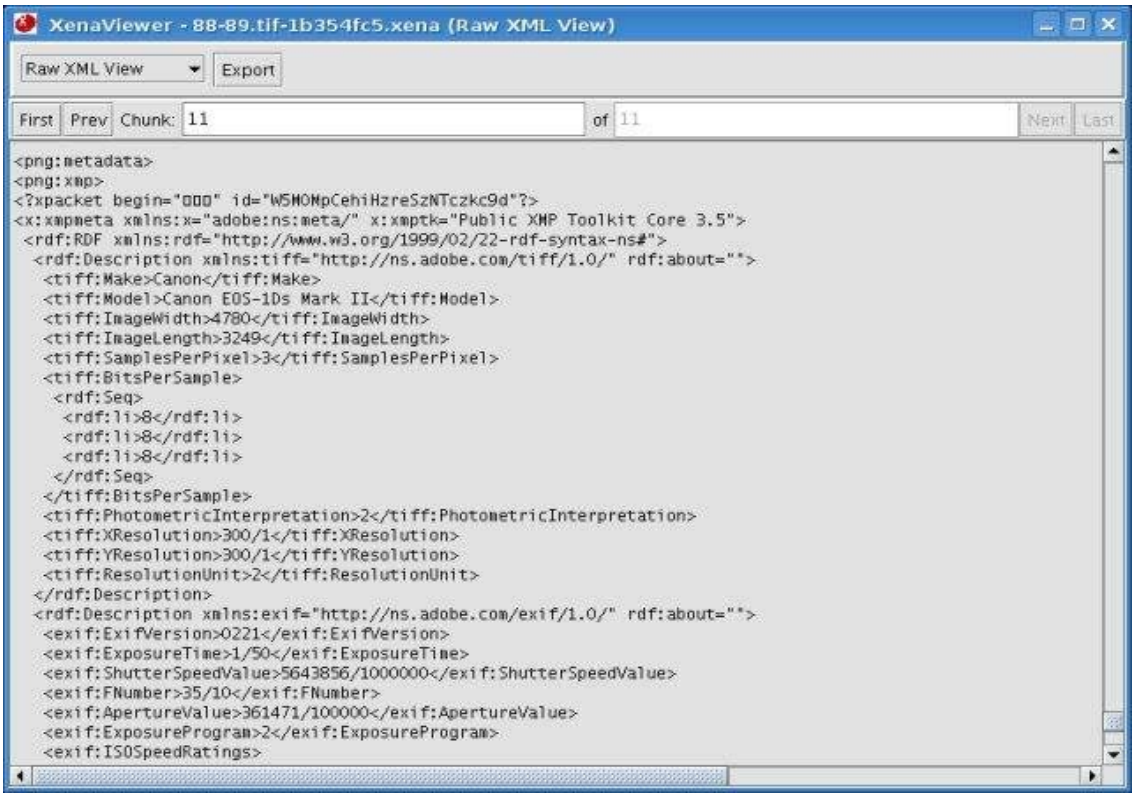
Comparativa de solucions comercials de preservació digital.

- **Ex-Libris Rosetta.** Sistema de preservació desenvolupat per la empresa israeliana "Ex Libris". Rosetta és un software que cobreix totes les necessitats de un sistema de preservació des de la fase d'ingesta a la de visualització. Està construït amb una arquitectura distribuïda la qual cosa el permet ser altament escalable. Basat en el estàndard OAIS, un dels punt forts d'aquest sistema és la existència d'una API que permet desenvolupar programari damunt aquesta plataforma, augmentant clarament el nivell de flexibilitat necessari. Té una interfície web per accedir a la seva configuració i disposa de plugins per customitzar el producte. Està àmpliament emprat per biblioteques i consorcis de biblioteques al voltant del món. Podem afirmar que és un dels millors sistemes que hi ha actualment al mercat. Té un preu alt i obliga a adquirir una llicència Oracle per al seu funcionament.



Imatge 1. Captura de pantalla del software Rosetta.

- **XENA (XML Electronic Normalising for Archives).** És un software de codi obert desenvolupat en Java per el Arxiu Nacional d'Austràlia. Està disponible gratuïtament baix llicència GNU. Entre les seves característiques destaquem:
    - S'executa tant a plataformes Windows com Linux
    - Posseeix plugins per identificar i convertir formats de fitxer.
    - API per llenguatge JAVA, per tal de desenvolupar programari a mida.
    - Processa arxius individuals o carpetes senceres mantenint la seva estructura.
    - Visor integrat per multitud de formats
    - Permet la execució en interfície gràfica o línia de comandes



Imatge 2. Visor de metadades de XENA

Externalització dels serveis de preservació

Una possibilitat per al nostre sistema de preservació és la de la externalització. Això és, cedir el nostre patrimoni cultural a una altre entitat que dugui la preservació del nostre fons i el d'altres institucions, centralitzant esforços i reduint costs econòmics. Aquesta opció és la més favorable als nostres interessos, però malauradament no hi ha cap projecte d'aquesta mena a la nostra comunitat, la qual cosa fa inviable la seva implantació. De totes maneres, la Universitat es proposa promoure un grup de feina amb altres institucions implicades per que, a mitjà termini, es pugui crear una supra-entitat que reuneixi els repositoris de les institucions que vulguin participar.

Solució triada

Degut als recursos disponibles per la universitat, s'ha determinat que la solució comercial propietària no s'ajusta al pressupost destinat per a la preservació, per això s'estudiarà la implantació del software XENA amb les pertinents modificacions. El personal tècnic de la universitat afirma que poden modificar i implantar un nou software basat en XENA gràcies a que el codi font d'aquest és de domini públic, gratuït, i amb llicència GNU. La solució passarà per modificar i desenvolupar plugins o petites aplicacions gràcies a la API Java que XENA ofereix.

Així doncs, la solució que s'implantarà al nostre sistema serà una adaptació de XENA, no és un fork del projecte. És una solució intermèdia entre un desenvolupament totalment propi i partint des de zero, i la solució comercial que és més cara però de més fàcil implantació.

Gràcies a que la Universitat disposa de recursos físics i de personal, la rectoria ha autoritzat el desenvolupament d'aquesta solució pròpia emprant personal actualment contractat per el departament de Centre de Tecnologies de la Universitat i, a més, la contractació d'una nova persona experta en temes de preservació digital per que pugui ajudar als actuals tècnics i responsables.

Amb el desenvolupament propi haurem de definir i implementar tota la estructura damunt la qual s'executarà el sistema de preservació. Veurem els detalls al apartat IV, magatzem de dades. També serà desenvolupament propi la web d'accés per als usuaris, tal i com comentarem al corresponent apartat.

Garantia del servei

Com a compromís amb els usuaris s'ha de definir un acord de nivell de servei on es pugui assegurar la màxima disponibilitat del sistema. En termes comercials això es diu un Service Level Agreement (SLA).

El sistema disposarà dels mitjans tècnics i humans per dur a terme una garantia de servei del 99,9% del temps, és a dir, s'accepten com a màxim 9 hores/any de servei aturat.

Aspectes legals

El nostre sistema emmagatzemarà i custodiarà les obres digitals d'acord amb les actuals legislacions en el tema. Aquestes són

- LOPD: llei de protecció de dades (LO 15/1999). Per a les dades de caràcter personal i sensible que en puguem emmagatzemar com ara llistats de persones, matrícules, qualificacions...
- Llei de propietat intel·lectual (RDL 1/1996). Per a la custòdia de les obres adquirides o publicades per altres editorials.

Anàlisi econòmic

El aspecte econòmic d'un projecte de preservació digital no és trivial. Intervenen moltes persones i es necessita una gran capacitat de magatzem així com de gestió de les dades, a més de les adquisicions que es facin.

En aquest apartat, sense entrar a valorar xifres concretes, es detallaran quins són els punts que es poden emportar la major part del pressupost.

En el pla econòmic del nostre projecte ha de figurar:

- Costs de personal. Com varem veure hi haurà personal de molts perfils diferents: tècnics, legals, administratius… Molts d'ells ja hi eren al actual sistema, però es preveuen un parells de contractacions de perfils amb experiència a preservació digital.
- Cost de desenvolupament de les noves eines.
- Costs de materials. Veurem que a la fase de Magatzem de dades es detallaran els discs, racks i servidors necessaris per a dur a terme la nostra tasca.
- Costs de adquisició de recursos a editorials i altres institucions. Aquest cost tampoc és nou, ja que la universitat adquireix material nou cada any, però en format analògic. Ara haurem de negociar si hi haurà un increment de preu per demanar una còpia digital a les editorials.
- Cost de la introducció de les dades dins el sistema. Representen el 50% del cost d'un projecte de preservació degut a que requereix molt de temps i personal i maquinari especialitzat.
- Costs de manteniment:

- Elèctric
- Refresc de suports
- Substitució d’altres components obsolets (escàners, ordinadors de consulta).

3.2. Administració i gestió del dipòsit digital

El segon apartat funcional del model OAIS, administració, durem a terme les tasques d'administració i gestió del dipòsit digital. Aquest punt es centra en la tasca de manteniment del sistema i en la millora dels processos, així com a donar suport als usuaris i relacions amb el exterior, com ara contacte amb altres institucions o repositoris i la adquisició de recursos de altres entitats o editorials.

Revisar procediments

Per a una millora contínua en els processos del nostre pla de preservació és molt important revisar els procediments que es duen a terme. En el nostre cas, el cap de biblioteca, juntament amb assessors tècnics i jurídics, revisaran els actuals procediments anualment per tal de detectar canvis que puguin afectar al nostre sistema. Alguns d'aquests canvis poden ser:

- Canvis legals: el material custodiat per el nostre sistema de preservació està subjecte a la actual normativa legal en quant a protecció de dades, LOPD, i a drets d'autor. Qualsevol canvi legislatiu en aquestes matèries poden afectar al nostre repositori, obligant a incorporar nou material o llevar-ne.
- Canvis en la tecnologia: el món de la tecnologia és ràpidament canviant. El nostre sistema de preservació digital es veu afectat per tres canvis tecnològics importants: canvis al tipus de magatzem de dades, canvis als sistemes on s'executa el software preservat, i canvis als formats dels fitxers. Per tot això anualment es revisaran prioritàriament aquest 3 punts per detectar qualque millora al procés, com ara, canvi de tecnologia d'emmagatzemament, o començar a emprar un nou format molt més avantatjós per al nostre propòsit de preservació, o per que el format que empraven ha quedat obsolet.
- Canvis pressupostaris: qualsevol canvi al pressupost destinat al sistema ens afecta considerablement. Si el pressupost augmenta, podem disposar de més recursos personals i hardware per a millorar els nostres processos, com ara, hardware més avançat o millors màquines digitalitzadores. Si el pressupost disminueix, haurem de plantejar-nos deixar de preservar certs recursos o disminuir les aportacions externes al sistema, com ara reduint les adquisicions.
- Refinament i actualització dels processos: tot i que no es detectés cap canvi extern, hem de fer revisions internes dels processos, com ara veure si hi ha alguna part del sistema que no funciona com havíem dissenyat, o si les nostres planificacions inicials del rendiment no s'ajusten a la realitat, normalment per optimistes. Això far que s'hagin d'actualitzar els processos del nostre pla de preservació i corregir aquestes errades de planificació inicial.

Suport als usuaris

Una part important de tot sistema amb usuaris externs, és la de donar suport a aquests. En el nostre cas, hem decidit que siguin els actuals bibliotecaris els qui donin suport als usuaris finals del sistema. Aquest usuaris accedeixen al sistema mitjançant un accés web que es desenvoluparà per permetre'ls fe cerques dins el sistema de preservació.

Un altre aspecte del suport dels usuaris és la de la gestió dels propis usuaris i els seus permisos. Disposarem d'un sistema de gestió d'usuaris on els donarem d'alta, gestionarem el seu password i els permisos d'accés al sistema.

- Alta: tot usuari de la comunitat universitària, personal docent, administratius i estudiants, tenen dret a tenir un usuari d'accés al sistema. Aquest usuari serà el document d'identitat de la persona.
- Password: inicialment els donarem un password que caducarà en 24h i l'hauran de canviar per un de la seva elecció. El personal de la biblioteca podrà tornar regenerar el password d'un usuari en cas de pèrdua.
- Permisos: hi haurà diferents perfils d'accés al sistema, depenent dels documents que s'hagin de consultar. Un usuari pot tenir més d'un permís.

Contacte exterior

La gestió de la preservació és una tasca complicada. Com veurem més endavant es necessita gran quantitat de recursos per dur-la a terme. Això fa que les institucions contactin unes amb les altres amb la finalitat de compartir recursos i no duplicar cert material.

Mitjançant acords amb altres biblioteques, podem estalviar-nos la preservació de certes obres del nostre catàleg, per que altres biblioteques ja les estan preservant i ens donen accés al seu sistema. Amb aquesta modalitat s'haurà de fer qualque tipus d'integració dels seus sistemes dins el nostres, ja que probablement no serà el mateix. Això fa que les consultes al apartat d'accés, també hagin de poder llençar consultes al sistemes externs amb els quals estem connectats.

Així mateix, i recíprocament, la nostra biblioteca també pot arribar a acords amb altres biblioteques per oferir els nostres materials, possibilitant que accedeixen als nostres sistemes de preservació.

Per tot això seria desitjable un sistema estandaritzat d'intercanvi de dades entre sistemes de preservació o, degut a la envergadura de la tasca, en un futur ajuntar tots els sistemes dins un sol mateix i que la resta d'institucions hi accedeixin.

Adquisició de recursos

És un cas particular de contacte amb l'exterior. La adquisició de recursos té més a veure amb comprar nou material a editorials o productores, que a la integració de sistemes informàtics. S'han d'establir polítiques d'adquisició sempre que hi hagi partida pressupostària assignada. Fins ara la universitat feia adquisicions a les principals editorials i revistes científiques. Amb la arribada dels documents digitals no hi ha d'haver molt de canvi en aquest aspecte, ja que la majoria de productores i editorials ja ofereixen el seu contingut també en format digital i no només físic (llibre/DVD…).

El recursos adquirits hauran de passar per la fase d'ingesta com tots els altres documents.

3.3. Ingesta

La fase d'ingesta és la fase d'introducció de recursos dins el sistema de preservació. Veurem quines metodologies emprarà el nostre sistema i les tecnologies emprades.

Digitalització

La digitalització és la tècnica mitjançant la qual incorporarem al sistema de preservació digital tot aquell fons documental actualment en format analògic. És un procés lent i costós. S'ha d'estudiar si fer-ho per a totes les obres o només per aquelles que tenen major valor de conservació. En el nostre cas, s'ha fet una selecció inicial d'obres de major rellevància per a la seva preservació per el seu interès i valor artístic i cultural. Tot i així, algunes obres es deixen en cua per a una posterior digitalització quan el temps i el pressupost ho permetin.

La digitalització fa servir maquinari molt específic (típicament escànners òptics en el cas de llibres) i personal especialitzat. A més, degut al caràcter únic d'algunes obres s'ha de fer amb molta cura, lo que fa que es dediqui un temps considerable per obra. Aquestes són les causes del seu elevat cost.

En el nostre cas, farem servir escànners òptics amb software de reconeixement òptic OCR, de tal forma que el producte digital resultat del procés, pugui ser indexat i tractat com un document de text digital i no com a imatges. Això ens facilitarà les cerques dins el contingut del document i facilitarà la feina dels usuaris. El procediment es fa a dues fases:

1. Fase d'escàner òptic: es posa la obra dins l'escàner i com a resultat obtenim una imatge digital. Nosaltres emprarem dos models diferents depenent del tipus d'obra:
    a. Obres úniques amb gran valor històric que necessiten d'una manipulació cuidadosa i lo menys intrusiva possible: *FUJITSU ScanSnap SV600*. Té una velocitat d'escaneig de una pàgina cada 3 segons i el seu preu de mercat ronda els 600€.

a. Obres on el contingut és lo més important i no el suport, model *FUJITSU fi-6400*. Té una capacitat d'escaneig de més de 100 pàgines per minut. El seu preu de mercat ronda els 9,000€.

1. Fase de reconeixement òptic de caràcters (OCR): un software reconeix els símbols com a caràcters i composa un document de text digital amb ells. Aquest fitxer serà el incorporat dins el sistema de preservació digital. S'ha de tenir curar amb aquest procés ja que, tot i que el software cada vegada és més avançat, es poden trobar errades de reconeixement. Emprarem el software que el fabricant dels escàners, en aquest cas *FUJITSU*, aporta amb cadascun dels models.

La política del nostre pla pel que fa a la preservació dels documents originals, continuarà essent la mateixa que fins ara. No destruirem el document original només pel fet que ara estigui digitalitzat, si no quan la política de preservació de llibres ho dictamini, ja sigui per deteriorament, falta de espai o pèrdua de valor intel·lectual.

Canvi de format

Una de les tècniques aplicades en la fase d'ingesta, és la de la modificació del format de fitxer si no es correspon amb algun dels definits com a idonis al pla de preservació.

Com varem veure en apartats anteriors, es varen definir una sèrie de formats preferits, acceptats i problemàtics per a diferents tipus de fitxers (veure Taula 1).

En el nostre sistema disposarem de diferents programaris dedicats a la conversió dels formats problemàtics en formats preferits o, al manco, en format acceptats, per tal de que la preservació estigui més estandaritzada.

Com a exemples, tindrem:

- Total Video Converter per a formats de vídeo http://www.effectmatrix.com/total-video-converter/
- AVS Image converter per a les imatges http://www.avs4you.com/AVS-Image-Converter.aspx
- AVS Document Converter per als fitxers de texte http://www.avs4you.com/AVS-Document-Converter.aspx

La utilització d'aquestes eines s'ha d'automatitzar per tal de que, en arribar un recurs digital que s'hagi de convertir, no es necessiti cap intervenció humana (o la mínima possible), ja que el volum de dades que s'han de convertir pot ser gran.

JHOVE (https://sourceforge.net/projects/jhove/) és una eina desenvolupada per la Universitat de Harvard, al Estats Units, que identifica automàticament el format del fitxer i ens indica si estan correctament formats d'acord amb la norma del format. Te la capacitat d'extreure gran quantitat de metadades tècniques.

Generació de metadades

Una de les tasques més importants de la fase d'Ingesta és la de la generació de les metadades associades al fitxer digital dins el sistema. Com varem veure les metadades són fonamentals per a un bon funcionament del sistema de preservació, per a la classificació i posterior cerca dels documents dins tot el fons digital.

Com varem veure al apartat anterior, hi ha una gran quantitat de metadades, de diferents tipus, que es poden extreure per als recursos digitals. De entre elles aquí en parlarem de les metadades de preservació, per ser les més importants per al funcionament del nostre sistema i per les particularitats que tenen.

Per al conjunt de metadades de preservació emprarem el format de PREMIS, degut a la seva extensió i definició com a estàndard de-facto.

Exemple de esquema PREMIS (objectCharacteristicsExtension):

```
<premis:mdSec>
 <premis:mdWrap MDTYPE="TEXTMD" MIMETYPE="text/xml">
```

```
        <premis:xmlData>
        <textmd:textMD xmlns:textmd="info:lc/xmlns/textMD-v3">
        <textmd:character_info>
        <textmd:charset>ISO-8859-1</textmd:charset>
        <textmd:byte_order>little</textmd:byte_order>
        <textmd:byte_size>8</textmd:byte_size>
        <textmd:character_size>1</textmd:character_size>
        <textmd:linebreak>CR/LF</textmd:linebreak>
        </textmd:character_info>
        <textmd:markup_basis version="1.0">
        XML</textmd:markup_basis>
        <textmd:markup_language>
        http://www.loc.gov/standards/alto/ns-v2</textmd:markup_language>
        </textmd:textMD>
        </premis:xmlData>
 </premis:mdWrap>
</premis:mdSec>
```

Un exemple d'us de PREMIS amb l'estàndard METS el trobam a la següent URL http://www.loc.gov/standards/premis/louis-2-0.xml , que no copiam aquí degut a la seva envergadura. En aquest exemple podem veure com les metadades PREMIS damunt un fitxer digital (en aquest cas una imatge de Louis Armstrong) queden descrites amb la codificació METS.

Comprovació d'integritat (*checksum*)

Una de les metadades més importants i fàcils de calcular d'un fitxer digital es la seva suma de verificació o checksum. Aquesta signatura, que és una cadena de lletres i números (hash), ens ajuda a comprovar que el fitxer no ha estat manipulat ni ha sofert cap alteració.

Hi ha diferents algoritmes de càlcul de la clau hash, com ara les funcions SHA-1, SHA-256 o MD-5. El nostre software emprarà SHA-1, ja que s'ha provat robust amb el pas dels anys i no se l'ha trobat cap vulnerabilitat. SHA-1 genera una clau hash de 20 bytes i té un límit màxim del missatge de 2^{64} bits.

Un exemple senzill de hash, emprant una cadena de text és:

```
sha1("Hello") = f7ff9e8b7bb2e09b70935a5d785e0cc5d9d0abf0
```

Aquesta funció és extensible a tota la cadena de bytes d'un recurs digital, que no reproduïm aquí per la seva extensió.

El sistema de preservació emmagatzemarà juntament amb el recurs digital, la seva clau hash SHA-1. Quan es torni a accedir al recurs (fase d'accés), es tornarà calcular la seva

clau hash SHA-1 i es comprovarà amb el valor emmagatzemat. Si els valors coincideixen és que el recurs no ha sofert cap alteració en el transcurs de la seva vida dins el sistema de preservació i podem afirmar que és el mateix recurs que es va introduir al sistema a la fase d'Ingesta.

S'ha de recalcar que la clau de hash no assegura autenticitat, només verifica que el fitxer no ha estat alterat.

Altres controls

Hi ha més controls apart dels descrits anteriorment, com ara el Control antivirus: especialment indicat en el cas de recepció de fitxers executables, però extensible a tots els fitxers rebuts siguin del tipus que siguin. S'emprarà la solució corporativa implantada als ordinadors de la universitat, Kaspersky Antivirus (www.kaspersky.es/Antivirus).

3.4. Magatzem de dades

En aquest apartat es detalles les solucions tècniques triades per a posar en marxa una de les parts més crítiques d'un sistema de preservació, el magatzem de dades.

Gestió del sistema de magatzem

El sistema de magatzem és el lloc on s'emmagatzemen físicament les dades. Està format per *racks* de discs durs magnètics d'alta fiabilitat controlats per protocols de còpia de seguretat i redundància.

La infraestructura estarà composta de tants discs durs com siguin necessaris per emmagatzemar tota la informació instal·lats en *racks*. Aquests racks seran específics per a preservació digital i duran el seu propi gestor d'energia o SAI per tal de mitigar les pujades i baixades de tensió i el renou en la corrent alterna que podrien fer malbé els discs.

Imatge 3. Rack de 4 discs SATA.

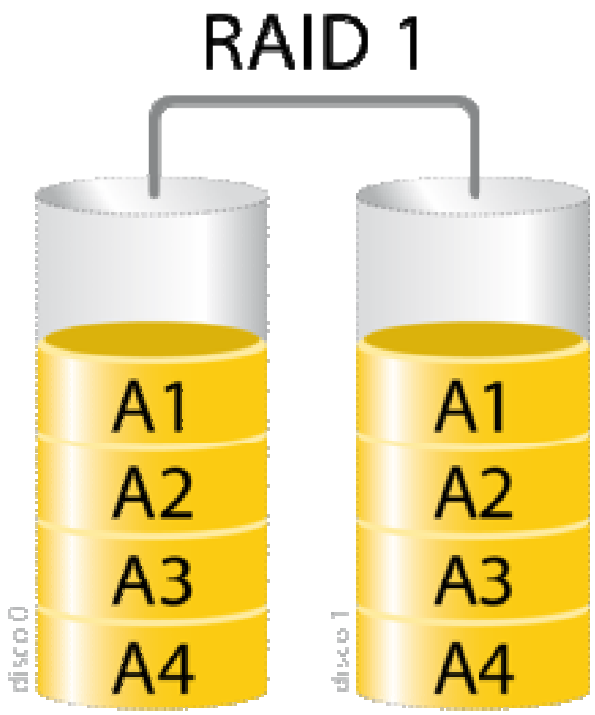
La tecnologia dels discs durs emprada serà interfície SATA-III (6GBps) que proporciona altes velocitats de transferència per tal de que el sistema tengui un bon rendiment. Els models de disc dur seran de mida de 2TB cadascun, de tecnologia magnètica (no SSD). Els discs SSD, tot i que tenen molt bon rendiment, son menys fiables i molt més cars. Encara no són una opció viable per al nostre sistema, però no descartam que ho puguin ser en el proper reemplaç de discs.

S'ha fet una valoració inicial de les necessitats d'emmagatzematge del nostre sistema i, amb el volum d'obres i degut a que la majoria són texts (llibres, projectes de fi de màster, treballs, blogs…), s'ha estimat una necessitat de 40TB. Això són 20 discs durs de 2TB en 5 racks de 4 discs.

Ara bé, s'han de tenir en compte les necessitats de replicació i de còpia de seguretat. En el nostre cas, tindrem un sistema mirall totalment replicat amb RAID 1 (mirroring).

RAID 1 fa una còpia exacta de les dades d'un disc a un altre. En el nostre cas emprarem una configuració de RAID 1 amb dos discs, però es poden afegir més. Aquesta configuració augmenta considerablement la fiabilitat del sistema de dades ja que la probabilitat de fallada del conjunt és igual al producte de la probabilitat de fallada de tots els discs (han de fallar tots dos a la vegada). Suposem que la probabilitat de fallada de un disc dur és de un 2% anual (0,02). La probabilitat de fallada d'un sistema RAID1 amb replicació de dos discs serà de 0,02 x 0,02 = 0,0004 (0,04%).

Com hem dit, aquesta solució suposa duplicar tota la infraestructura, duplicant costs, però donat que les nostres necessitats no són especialment elevades (20 discs durs inicials és pot considerar un sistema petit), els tècnics de la Universitat aconsellen emprar aquesta tecnologia per la seva facilitat de configuració i administració. Per exemple, una d'aquestes facilitats d'administració la podem trobar si un dels discs fallés: encara quedaria el segon funcionant, i substituint el disc defectuós per un nou i copiant les dades d'un al altre, tendríem una altra vegada el sistema funcionant, amb redundància, sense aturada del servei.



Imatge 4. Mirroring de RAID1.

RAID1 pot millorar la velocitat de lectura ja que un mateix recurs es pot llegir des de dos discos diferents a la vegada, ja que està replicat. Però no millorarà la velocitat d'escriptura ja que una mateixa dada s'ha d'escriure a tots dos discos alhora.

Hi ha controladores RAID1 que disposen d'un mode d'operació que ens pot beneficiar bastant a la hora de verificar les dades: fer lectures de la dada simultàniament des de

els dos disc i fer comprovacions per detectar errades. Amb aquesta configuració guanyaríem en integritat de dades però no en velocitat, que seria la del un sol disc.

Així doncs, les necessitats d'espai es dupliquen i queda en un total de:

- 40 discs durs de 2TB
- 10 racks de 4 discs.
- 80TB d'informació total.

S'ha de tenir en compte que normalment es deixa un 15% de la capacitat dels discs buida, per tal de no afectar al rendiment d'escriptura/lectura. Aquest càlcul ja s'ha tingut en compte en calcular les necessitats inicials.

En quan al creixement de la infraestructura d'emmagatzematge, es calcula que anualment es necessitaran 4TB addicionals (més 4 de la rèplica), lo que suposa un 10% anual aproximadament, o traduït en discs, un increment de 4 discs durs de 2TB anualment. Aquest cost és perfectament assumible per la organització que dona el vistiplau a la implantació d'aquesta infraestructura.

Un dels desavantatges de la configuració RAID 1 és que cada vegada que vulguem ampliar el nostre sistema de magatzem digital, s'ha d'incloure un número parell de discs (original i mirall), fent que els costs siguin sempre el doble. Mentre les necessitats de emmagatzematge estiguin controlades i no es disparin (hem parlat de 4 discs/any aproximadament), aquest sistema pot funcionar be. Si en parléssim de dotzenes de discs nous cada any, per ventura hauríem de plantejar un altre sistema RAID, com ara RAID5, que fa ús de blocs de paritat i no duplica tota la infraestructura de dades, només afegeix un tant percent de redundància.

Còpies de seguretat

S'ha de dir que el sistema RAID no és un sistema per còpies de seguretat, si no un sistema de replica de dades per assegurar la seva disponibilitat si un disc deixa de funcionar. El sistema RAID no assegura les dades de modificacions (intencionades o no). A més no assegura la supervivència de les dades davant un incident al centre de dades, ja que els discs rèpliques són a la mateixa ubicació que els originals.

Per tot això es faran còpies de seguretat a suports de cinta magnètica de tipus Ultrium LTO-7, amb una capacitat de 6TB (i fins a 15 TB comprimits). Aquestes còpies s'emmagatzemaran a una altra ubicació geogràfica, un altre edifici dins el campus de la universitat, per tal de complir amb els estàndards de còpies de seguretat de la Universitat.

Imatge 5. Cintes magnètiques LTO-7

Hi ha varis fabricants que presenten models LTO-7 al seu catàleg. La Universitat fa feina amb el model *FujiFilm LTO 7 Tape with Barium Ferrite (BaFe)*, que tenen un preu estimat de mercat d'uns 120€ per unitat. Es calcula que en necessitarem unes 6.

La política de còpies de seguretat definida serà la següent:

- Còpia completa de totes les dades els dilluns a la 1 de la matinada. Es calcula que el procés pot trigar entre 3 - 4h degut a les altes tasses de transferència de les cintes LTO-7 (fins a 2,5TB/h)
- Còpies diferencials cada dia de matinada, de dilluns a divendres, a la 1.

Així mateix, tot el centre de dades comptarà a un punt de suministrament elèctric propi i mecanismes de alimentació elèctrica ininterrompuda (o SAI) de la suficient capacitat com per poder disposar d'uns 10 minuts de servei en cas de fallada elèctrica.

Refresc de suports

Com sabem, els suports físics on emmagatzemam les dades (en el nostre cas els discs durs de 2TB), no tenen una vida il·limitada. Encara que hagin sigut tractats amb les millors condicions i preservats en òptimes condicions de temperatura i humitat, arribarà un moment en el que fallaran. També es pot donar el cas que, en un canvi de tecnologia, acabin per quedar obsolets i no els podem fer servir més per impossibilitat de connexió (com per exemple quan es va canviar la interfície de IDE a SATA). Per aquestes raons hem de tenir preparat un pla de refresc de suports.

El pla refresc de suports estudia la vida útil dels aparells i no espera a que deixin de funcionar. En comptes d'això s'estableixen polítiques de reemplaç dels dispositius mentre aquests encara funcionen. Tot i que amb una configuració RAID1 de replicació en mirall i una adequada política de còpies de seguretat la fallada d'un disc no

representa una amenaça greu, la obsolescència tecnològica dels discs sí que ens pots suposar una amenaça, ja que afectaria a la totalitat del nostre magatzem.

Per aquests motius s'estableix una política de refresc de suports que no és gaire restrictiva, i es basa en:

- Estudiar els canvis tecnològics en suports d'emmagatzematge digital: els canvis tecnològics o la obsolescència en els suports de magatzem no es sol fer patent amb molta freqüència. De fet la tecnologia de discs durs magnètics porta amb nosaltres més de 30 anys i la interface de connexió SATA (encara que ha anat evolucionant és compatible amb les anteriors) té ja més de 10 anys. Per això no es preveuen grans canvis en un futur immediat a part del canvi de tecnologia dels discs magnètics cap a SSD. Cada any es revisarà l'estat actual de la tecnologia i el preu dels dispositius per comprovar si mereix la pena canviar cap a SSD o no.
- Estudiar els paràmetres de vida útil del actual hardware (discs, interface, racks...): aquesta línia d'actuació es basa es estudiar els paràmetres S.M.A.R.T. dels discs durs on es pot preveure la fallada dels discs. Aquests paràmetres avaluen diferents estats dels discs i els valoren dins una escala. Si qualque valor sobrepassa un valor màxim es llença una alerta. Normalment es solen associar a una degradació del dispositiu.

ID	Name	Value	Threshold	Worst
01	Raw Read Error Rate	100	51	100
02	Throughput Performance	252	0	252
03	Spin Up Time	90	0	90
04	Start Stop Count	97	0	97
05	Reallocated Sector Ct	252	0	252
07	Seek Error Rate	252	0	252
08	Seek Time Performance	252	0	252
09	Power On Hours	100	0	100
0A	Spin Retry Count	252	0	252
0B	Calibration Retry Count	99	0	99
0C	Power_Cycle_Count	97	0	97

Imatge 6. Paràmetres SMART

3.5. Gestió de les dades

En aquest bloc es mantenen, sobretot, les metadades obtingudes a la fase d'ingesta. El seu correcte manteniment és clau per a la supervivència del sistema de preservació. També es tracten altres temes com el registre de canvis (tots els canvis pels que ha passat un determinat recurs digital des de que és dins el nostre sistema i fins la data actual) o la migració de formats, entre altres.

Gestió de les metadades

Les metadades extretes a la fase d'ingesta es gestionen en aquest bloc. Com varem veure al anàlisi teòric, hi ha dos modes de gestionar aquestes metadades:

- Juntament amb el recurs digital que descriuen
- Com a dades independents

Ja varem veure els pros i contres de aquests dos tipus de gestió, i en el nostre cas ens hem decidit per gestionar les metadades com a dades independents però mantenint les metadades dins el fitxer també, és a dir, un sistema mixt. Això és per que tindrem major control de les mateixes i afavorirà la seva consulta.

El nostre sistema de magatzem de dades tindrà un espai reservat per als fitxers XML amb les metadades generades per la fase d'ingesta en format PREMIS i METS i els tractarà també dins el pla de preservació. D'aquest mode cada recurs digital portarà dins de si mateix la informació necessària per a la seva identificació i utilització i, al mateix temps, els motors de cerca dels recursos de la fase d'accés agilitzaran la seva cerca sense entrar en contacte directe amb el recurs.

Aquest model mixt de gestió de les metadades té un problema associat a la replicació de les mateixes: en podem tenir dos versions diferents de metadades per a un mateix recurs (les del propi recurs i les gestionades externament). Per això és obligat que el nostre sistema dugui un control exhaustiu dels canvis dins els recursos i actualitzi immediatament totes les metadades associades a aquest.

Físicament les metadades s'emmagatzemaran de forma separada però de dins el nostre magatzem de dades.

Migració de formats

Com ja hem parlat abans a la fase d'ingesta, es controlen el formats d'entrada dins el nostre sistema de preservació. Però una vegada dins el sistema, els recursos també

poden sofrir canvis als formats, principalment per que hagin quedat obsolets amb el pas del temps, encara que no és la única raó.

El nostre pla de preservació ha de fer un seguiment dels formats emprats i actualitzar el document definit al bloc de Planificació de la preservació on es defineixen quins són els formats preferits, admesos o problemàtics. Si un format preferit passa es rebaixa de categoria a admès o problemàtic, seria convenient transformar tots els fitxers d'aquest tipus que estiguin dins el sistema de preservació cap al nou format.

Aquesta revisió es farà en la sessió de control anual al sistema i no està exempta de perills:

- S'ha d'avaluar que no es perden dades amb la conversió al nou format.
- S'ha de revisar que el nou document te les mateixes propietats significatives.
- S'han de revisar els programes visualitzadors per tal de verificar que són compatibles amb el nou format i mostren correctament el nou document.

Aquests riscos, juntament amb el fet que una migració de formats suposa un gran esforç tècnic, aconsellen que les migracions de formats es facin lo menys possible. S'ha d'intentar triar formats amb una llarga vida útil i fer els mínims canvis possibles en aquest aspecte, tot i que hi ha moments en que es farà inevitable.

Així mateix, qualsevol canvi en el format del document ha de quedar reflectit a un registre de canvis del que en parlarem més endavant.

D'altra banda el nostre pla de preservació ha de definir el destí dels recursos originals després d'haver fet la migració de formats. Hi ha dues possibilitats:

- mantenir el document dins el sistema juntament amb la nova còpia amb nou format
- eliminar la còpia amb el format problemàtic i només mantenir la nova

Degut als possibles riscos que comporta la migració de formats i a la relativament baixa capacitat de dades necessària, es triarà en un primer moment per la primera solució i mantenir ambdues còpies. A la fase d'accés, l'usuari podrà triar descarregar qualsevol dels formats que tenim emmagatzemats per a la seva visualització.

Una altra qüestió a tenir en compte és que, degut al gran esforç tècnic i en recursos que suposa la migració de milers de fitxers d'un format a un altre, si ens interessa fer-ho de cop per a tots els fitxers o millor a la fase d'accés quan el recurs es vulgui consultar per part d'un usuari. En el nostre sistema triarem la primera opció per que, encara que és més costosa, es farà molt de tant en tant.

Registre de canvis

Durant la vida d'un recurs digital dins el nostre sistema de preservació, aquest pot sofrir variacions, com ara la migració de formats que acabem de veure a l'apartat anterior. Això ens obliga a comptar amb un registre detallat dels canvis soferts per el recurs digital i documentar tot el procés.

Totes les accions que al llarg del temps es vagin executant damunt un determinat recurs digital hauran de quedar enregistrades mitjançant metadades PREMIS. Aquest informació ens permetrà seguir la traçabilitat del recurs i justificar els seus canvis de format i assegurar, davant tercers, la autenticitat del seu contingut. A més ens possibilita trobar possibles errades comeses al procés de migració.

Auditories externes

El nostre pla de preservació comptarà amb la realització anual d'auditories externes per tal de verificar la correcta gestió del magatzem de dades i de la gestió de les metadades.

3.6. Accés

El darrer apartat del que parlarem és l'accés a les dades. És un apartat fonamental del nostre sistema, sense el qual, tota la feina de preservació feta fins ara no hauria servir per a res. S'ha d'aportar una manera senzilla d'accés a les dades preservades a tots els usuaris amb accés al sistema. Això ens marca que hi haurà una interfície de consulta de dades i un control d'accés per tal que les nostres dades quedin visibles només a les persones que tinguin permisos.

Control d'accés

Un control d'accés efectiu a un sistema d'informació és fonamental per tal d'assegurar-nos que només les persones autoritzades accedeixen a les nostres dades. El sistema de control d'accés tindrà dues formes de fer login:

- Amb usuari-password donat per els serveis administratius de la biblioteca
- Amb una targeta identificativa temporal per a visitants

La primera manera serà la identificació més comuna. Tot el personal de la comunitat universitària disposarà dels seus credencials per accedir al material enregistrat dins el sistema de preservació. Aquest credencials es notificaran personalment la primera vegada que l'usuari vulgui accedir al sistema i li seran donats per el personal dels serveis administratius de la biblioteca.

La segona manera farà servir una targeta d'identificació temporal que s'insertarà a l'ordinador de la biblioteca. Aquest mètode es reserva per a visitants, investigadors o estudiants d'altres universitats amb les que s'hagi firmat un conveni, o en general qualsevol altra persona que la direcció de la biblioteca autoritzi. Les targetes no poden sortir del recinte de la biblioteca i s'han de tornar quan ja no es facin servir.

Les credencials atorgades caduquen a cada curs acadèmic i es renoven en començar el curs. Així limitam l'accés només al actual personal de la comunitat universitària.

Val a dir que no tots els usuaris tenen accés complet a totes les obres del catàleg digital. S'estableixen diferents rols (o permisos) per tal de limitar l'accés de les obres amb condicionants legals o de visibilitat.

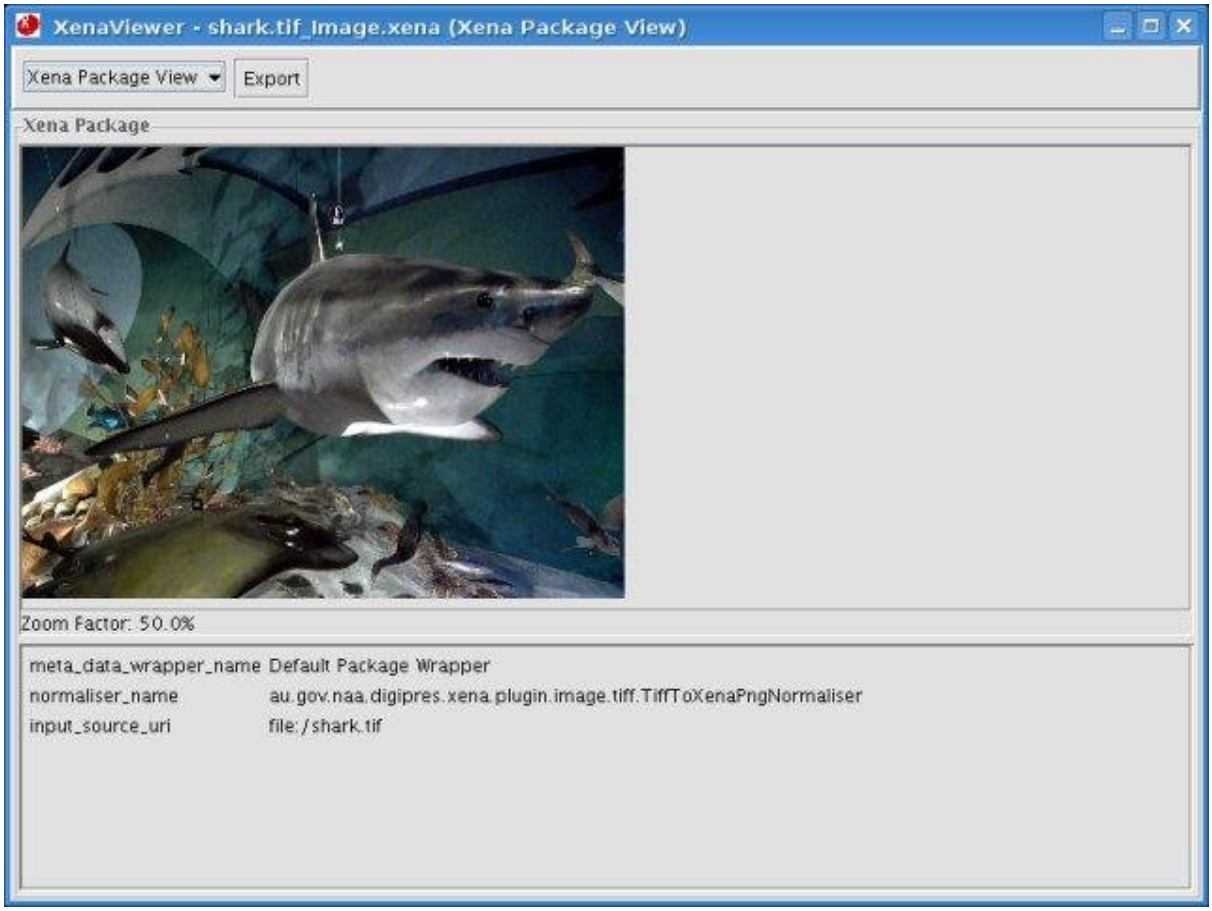
- Rol Estudiant: té accés als documents més comuns, com ara llibres, documentals o projectes d'altres companys
- Rol Investigador: té uns privilegis majors i té accés a certes bases de dades documentals restringides per temàtiques i a llibres de especial interès.
- Rol Accés Total: no hi ha cap restricció en el material al que pot accedir.

Interfície de consulta i accés

S'ha de preparar una interfície de consulta per a l'accés als documents digitals emmagatzemats dins el nostre repositori. Aquesta interfície primer de tot demanarà els credencials per a accedir al sistema.

Una vegada autenticat l'usuari, es mostrarà una pantalla amb el catàleg d'obres classificades per temàtica i un cercador d'obres. Fins aquí res diferent a lo que normalment és un cercador d'obres d'una biblioteca.

La diferència real d'un autèntic sistema de preservació digital i un simple cercador d'obres ve quan es tria la obra. Als antic cercadors lo màxim que es podia trobar són referències a les obres, les seves dades (metadades), i on trobar l'obra, però no consultar-la en línia. Amb el nou sistema de preservació digital, totes les obres estaran dins el sistema i consultables digitalment des de la aplicació.



Imatge 7. Visor del software de preservació XENA.

Com varem veure abans, una mateixa obra es pot presentar en diferents formats. Mostrarem al usuari totes les alternatives i proveirem d'un visor per a tots els formats acceptats per el sistema, mostrant també les seves metadades associades, tant les

tècniques com les de preservació. Aquest visor serà la peça fonamental de la interfície de consulta i estarà desenvolupat per el personal tècnic de la Universitat.

La descàrrega dels recursos no estarà disponible degut a la política de drets de les obres.

Emulació

La emulació és la tècnica mitjançant la qual mostram al usuari el recurs digital quan aquest ha quedat en un format obsolet o el software que servia per accedir-hi ja no es pot executar amb les màquines actuals.

El sistema visor de recursos del que hem parlat abans tindrà accés a diferents emuladors de sistemes obsolets amb el programari per visualitzar el document instal·lat.

Hi tenim dues maneres de fer la emulació:

- Amb un programa emulador vertader que s'hagi compilat per el nostre sistema i executi codi d'un sistema antic
- Amb la ajuda d'eines de virtualització

El primer cas, un programa emulador compilat per el nostre sistema és el que probablement ofereixi millor rendiment però a un cost molt alt. En el millor dels casos en trobarem un emulador que puguem emprar, ja sigui lliure o de pagament. Però en molts de casos s'haurà de escriure des de zero, amb l'elevat cost de temps i personal que això comporta. No és una alternativa viable per al nostre sistema de preservació per al recursos que tenim. Per això optarem per emprar eines de virtualització, molt més flexibles i ràpides.

Així doncs la solució triada serà emprar la virtualització com a tècnica per emular software antic i obsolet dins els nostres sistemes actuals. Aquesta tècnica ens permet oferir una àmplia varietat de sistemes a un cost reduït, tant econòmic com temporal. Per funcionar necessitam:

- Que el nostre hardware suporti la virtualització. Des de ja fa bastants anys els processadors ja inclouen aquesta tecnologia (Intel VT, AMD-V).
- Un software de virtualització, com ara Oracle Virtual Box, o VMWare. Aquest software s'executa dins el nostre sistema (host) i permet la execució d'un altre sistema operatiu.
- Una imatge o instal·lador del sistema operatiu original. Aquest sistema operatiu serà on s'executarà el software obsolet. Hem d'assegurar que el software

obsolet és compatible amb el sistema que estem instal·lant. Per exemple, per executar WordPerfect 6.1, instal·larem una imatge de Windows 95 a la màquina virtual.

- El software obsolet que ja no s'executa dins el nostre sistema, però que sí ho farà al sistema antic que estem emulant mitjançant virtualització. Per exemple, el WordPerfect 6.1.

D'aquesta manera tenim un entorn d'execució de software que, en principi, no era compatible amb els sistemes actuals, però que podem tornar a executar dins un contenidor com és la màquina virtual.

Aquesta solució és molt més ràpida d'implementar i a un cost relativament baix en comparació amb el desenvolupament d'un emulador propi. I és una solució més eficient que disposar de màquines amb tots els sistemes operatius que es preveu emprar. Amb un sol tipus de sistema amfitrió podrem executar gran varietat de sistemes emulats.

Imatge 8. Varis sistemes operatius executant damunt una distribució Linux.

S'ha de dir, però, que no tot són avantatges. La virtualització té un cost en tema de rendiment. Normalment el software que s'està executant dins una màquina virtual està penalitzat per la gran complexitat que agregam en temps d'execució. Però

normalment, aquesta penalització en rendiment és assumible ja que aconseguim la execució d'un codi que d'altra banda hagués estat molt difícil.

Annex. Motius per dur a terme un refresc de suports.

En aquest tercer i darrer apartat del projecte es vol estudiar la tecnologia darrera els suports físics digitals i, més concretament els motius per els quals s'han de reemplaçar.

Estructurarem aquest apartat en torn a aquests dos temes: deteriorament i obsolescència.

Introducció

Un dels temes que hem vist durant l'estudi de la Preservació Digital és el deteriorament i obsolescència dels suports on emmagatzemam la informació, en el nostre cas, discs durs. El deteriorament del suport es pot produir per:

- Deficients condicions de magatzem
- Mala qualitat del producte
- Mal ús per part del personal encarregat
- Factors externs, com ara pujades de tensió, inundacions...
- Degradació temporal dels components

Així mateix, tot i que aprofitéssim al màxim la vida útil d'aquests components ens enfrontem a un perill sempre present a la tecnologia digital: la obsolescència. Els productes van avançant, adquireixen millors prestacions i per suportar-les moltes vegades canvien els formats de connexió, les especificacions elèctriques i tota una sèrie de components que fan que els nostres components, encara que funcionals, deixin de estar suportats per el hardware actual i s'hagin de reemplaçar.

Tot això fa que no quedi més remei que substituir els suports antics per uns nous en lo que hem vist que és el refresc de suports per tal de no perdre les dades emmagatzemades en el dispositiu. Triar la millor tecnologia, una duradera, millorar els factors ambientals i tenir curar dels dispositius seran factors determinants per tal d'estalviar temps i recursos econòmics que es fan servir quan es refresquen el suports.

Deteriorament dels suports digitals

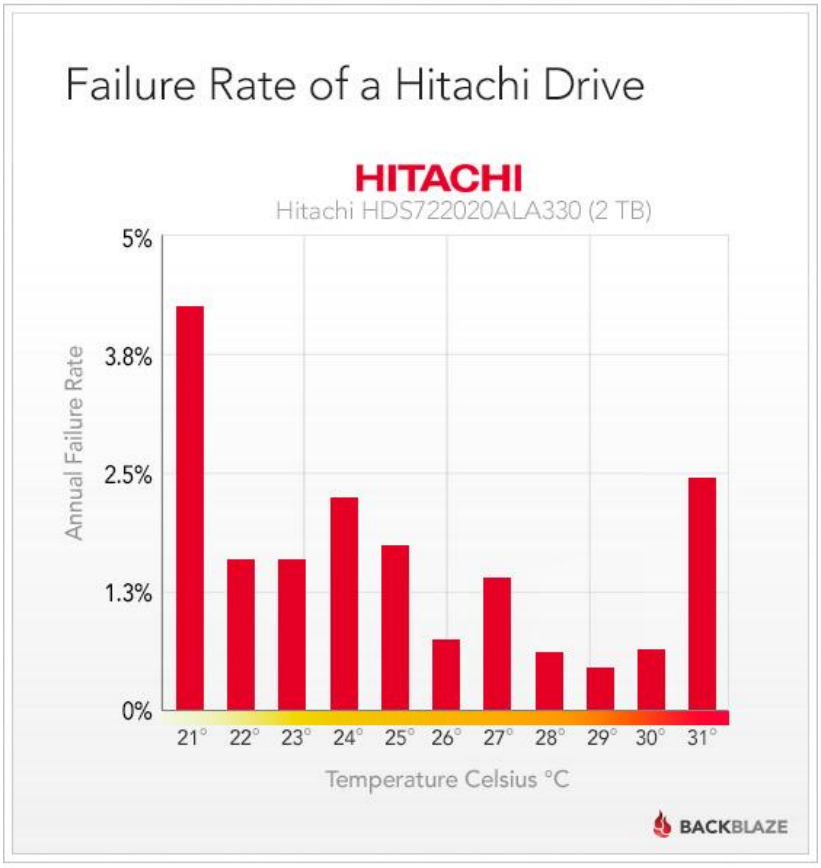
Estudiarem els diferents modes de deteriorament dels suports digitals abans esmentats i intentarem donar consells i solucions a cadascun d'aquests problemes.

Condicions de magatzem

Com a components físics, els discs durs necessiten d'unes condicions específiques per operar ja que, fora d'aquestes condicions es poden produir errades, o dit d'un altre mode, augmenta la probabilitat de fallada en el seu funcionament (rotació, lectura o escritura), i la seva vida útil es veu reduïda.

Els fabricants de discs durs recomanen unes condicions òptimes de magatzem per al seus dispositius. Aquestes són, aproximadament:

- Temperatura entre 25 i 30ºC
- Humitat entre 40 i 60%.



Imatge 9. Tassa de errades segons temperatura d'un disc dur Hitachi

Recents estudis confirmen que una humitat excessiva és molt més perjudicial per als sistemes de magatzem que una temperatura elevada. Si be tots dos són perjudicials, es recomana que la humitat estigui molt més controlada dins el nostre data center que la temperatura.

Les recomanacions que es fan en aquests aspectes són:

- Monitoritzar i mantenir temperatura i humitat dins els paràmetres indicats
- Vigilar més de prop els discs de més de 3 anys, ja que són més susceptibles a aquests canvis
- Evitar canvis ràpids dels paràmetres d'humitat i temperatura
- Separar els discs uns dels altres si hi ha possibilitat d'espai
- Posar els discs a la part de darrera del servidor permet una millor regulació de la humitat que els hi arriba.

Qualitat del producte

La qualitat dels diversos components que conformen un disc dur són molt importants quan es parla d'allargar la seva vida útil. Els discs durs són sistemes d'emmagatzematge magnètic i per funcionar fan servir plats amb un recobriment químic que reacciona al lector quan es posa damunt. La qualitat d'aquest recobriment farà que les errades a nivell de bit siguin majors o menors.

Un altre component crític és el component mecànic del disc. Per funcionar té un motor que mou una agulla lectora. Un mal funcionament d'aquest component pot espatllar diversos sectors del disc o tot el disc sencer.

En parlarem més d'aquests aspectes a la degradació temporal dels dispositius, però tenint en compte que discs de baixa qualitat degradaran abans que discs de millor qualitat.

La recomanació en aquest apartat és clara: apostar per material de qualitat, de fabricants solvents i que presentin dades contrastades de vida mitjana.

Ús dels dispositius

Un ús correcte dels dispositius és molt important per allargar la seva vida útil. Hi ha dues actuacions possibles en aquest aspecte:

- Ús humà, o manipulació dels discs. Encara que no serà freqüent manipular els discs una vegada instal·lats als racks, si per qualque motiu es necessita fer, s'han d'evitar les descàrregues d'electricitat electrostàtica. Per això hi ha unes bandes que eviten aquest tipus de descàrregues connectant un extrem a terra.

Imatge 10. Banda antiestàtica per evitar ESD (Electrostatic discharge)

- Ús per part del programari.
    - Moltes lectures i escriptures innecessàries
    - Posicionar la agulla en mode descans moltes vegades (Load_Cycle_Count)
    - Dades molt fragmentades que fan moure la agulla moltes vegades

Factors externs

Hi ha tota una sèrie d'amenaces i riscos externs que posen en perill en bon funcionament dels dispositius d'emmagatzematge. En podem llistar els següents:

- Pujades o caigudes de tensió: poden fer malbé els components electrònics dels discs i els seus controladors. Són més perilloses les pujades que els talls, però s'han d'evitar tots dos. Per mitigar els seus efectes es fan servir dispositius SAI i protectors de corrent. Hi ha diferents tipus de SAI, de més bàsic a més complet: SAI offline, SAI interactiu, SAI online.

Imatge 11. SAI Online. Transforma la energia alterna en contínua i altra vegada en alterna per netejar renou i variacions indesitjades.

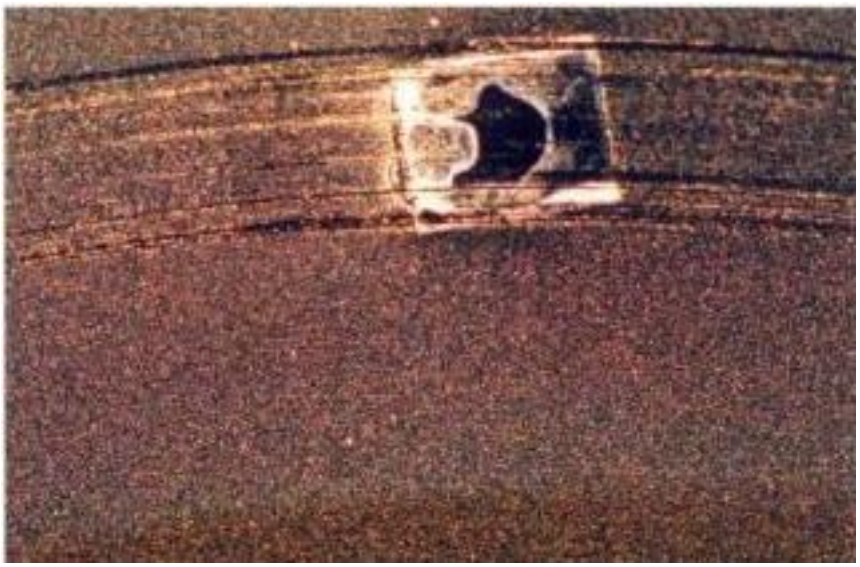
- Proximitat de camps magnètics. La proximitat d'aquests camps pot esborrar completament la informació que conté el disc.
- Inundacions o incendis. S'ha de poder establir un pla de prevenció de riscs i de mitigació d'aquests si apareixen. Escollir una bona ubicació per als dispositius és un bon començament. Disposar d'equips d'extinció o armaris ignífugs ajudarà en cas d'incendi.

Degradació temporal dels components

Amb el temps, els components que formen els discs es degraden i augmenten la taxa d'errades en les operacions de lectura i escriptura, i la probabilitat d'errada en qualque component del disc.

Dividirem els tipus de degradacions segons el component al que afecti:

- Degradació dels plats. El component principal del disc dur, és el plat metàl·lic on s'emmagatzemen les dades en forma de càrregues elèctriques gràcies a uns recobriments químics que reaccionen amb el magnetisme de la capçal. La humitat i la temperatura poden alterar les propietat ferromagnètiques dels plats, fent que les dades no s'escriguin ni llegeixin correctament.
- Degradació de la mecànica. El disc dur té dos components mecànics:
    - el motor de gir dels plats: far girar els plats a una velocitat constant de 5400 rpm o 7200 rpm. Una variació en aquesta velocitat farà que no s'accedeixi correctament a les dades.
    - el capçal que es mou a poques micres dels plats. Una errada en el seu funcionament pot fer que faci contacte amb els plats, ratllant-ho impedint l'accés als sectors afectats.



Imatge 12. Plat afectat per capçal

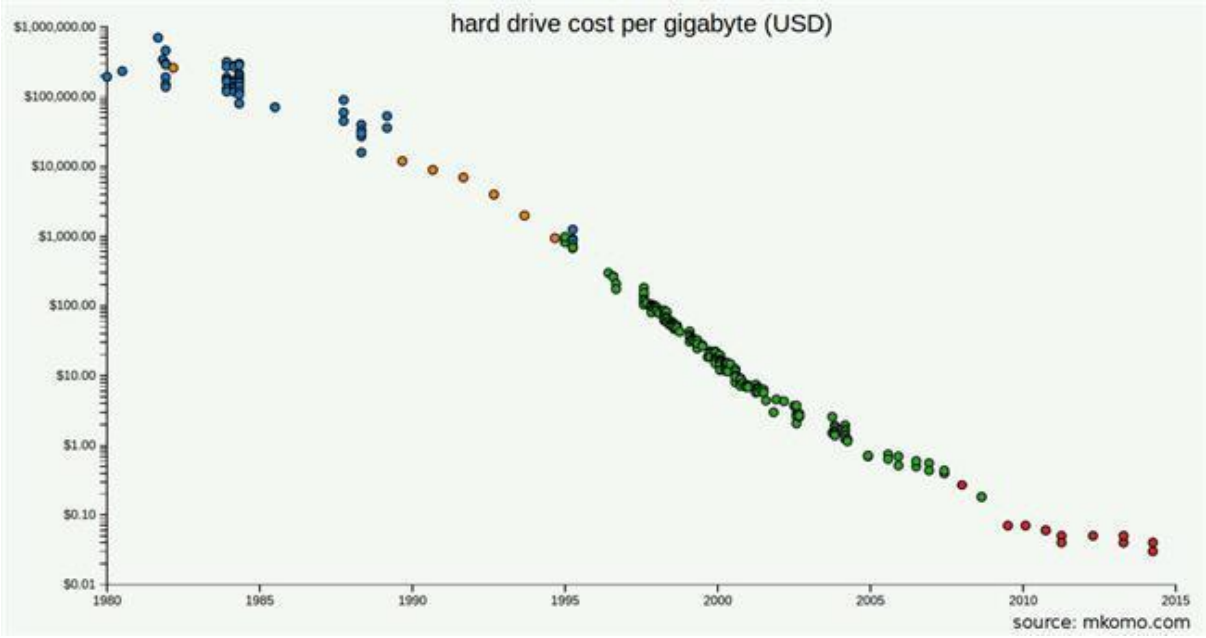
- Degradació de la circuiteria (controladors). La humitat pot degradar els components electrònics que controlen l'accés a les dades.

Evolució de la tecnologia i obsolescència

Vivim en un món on la tecnologia avança ràpidament. Això té conseqüències positives i negatives. Positives, es milloren les prestacions amb rapidesa. Cada any els components són un tant percent millors que l'any passat. Però negativament, moltes vegades per aconseguir aquestes noves prestacions s'ha de fer a canvi de modificar la tecnologia i deixar de donar suport a tecnologia més antiga que encara pot funcionar. Diem doncs que aquesta tecnologia ha quedat obsoleta, no per que no funcioni, si no per que el sistema al que va connectat, ha avançat tant que ja no es pot fer servir en ell.

Un clar exemple són els disquettes de 1,44MB. Va ser una tecnologia molt emprada fa 20 anys, però ja en fa temps que els ordinadors no inclouen un lector per aquests dispositius. Així, encara que podrien funcionar perfectament avui en dia, es considera tecnologia obsoleta per que ningú la empra, per haver estar sobrepassada per altres tecnologies que permeten més dades en un volum més petit (avui en dia podem trobar pen drives de 128GB, i tan sols ocupen uns pocs centímetres). Així i tot, encara avui es podria trobar un lector per disquettes i transferir les dades que tinguéssim emmagatzemades a un nou dispositiu (refresc del suport).

La evolució dels dispositius de magatzem de dades continua, any rere any, disminuint el preu per bit. Així, si fa 20 anys un GB costava devers 1.000$, avui en costa uns pocs cèntims.

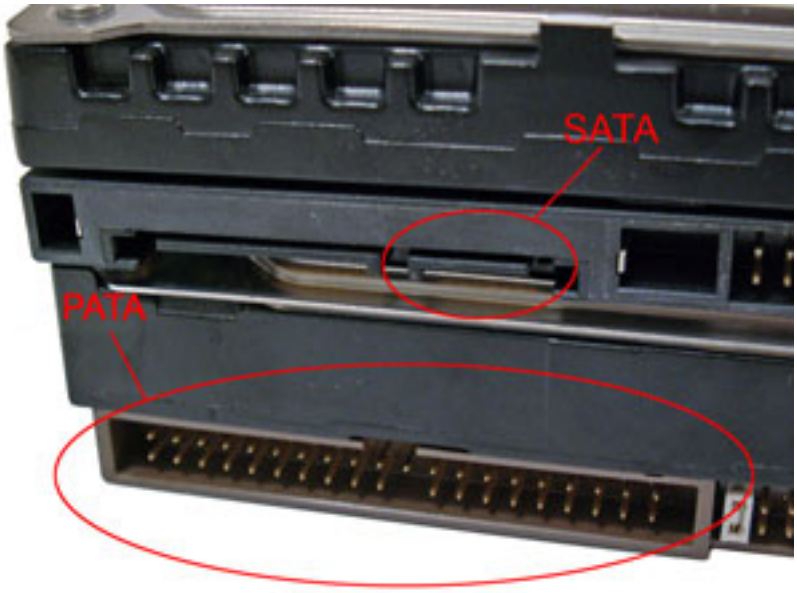


Imatge 13. Preu per Gigabyte en dòlars americans. La escala vertical és exponencial.

Afortunadament, en el món del emmagatzematge digital, els canvis de tecnologia succeeixen molt poc a poc. Encara fem feina amb disc magnètics, com fa 30 anys, tot i que la densitat ha millorat moltíssim, seguim amb la mateixa tecnologia. Només en aquesta dècada es comença a veure els pròxims substituts de la tecnologia magnètica, els disc d'estat sòlid o SSD que no tenen parts mecàniques ni magnètiques, són tot circuiteria.

Els darrers canvis que afecten a la obsolescència de discs antics s'han produït als tipus de connexió del disc amb la placa. Passant de P-ATA (paral·lel) cap a S-ATA (serial), bàsicament per permetre transferències de dades més ràpides. Les taxes de transferència de PATA anaves dels 16MB/s fins a els 166MB/s, mentre que SATA arriba actualment fins als 600MB/s.

El estàndard P-ATA va durar des de 1986 fins al 2004, moment que va començar a ser reemplaçat progressivament per el S-ATA. Però encara avui venen controladores de disc PCI per connectar discs P-ATA.



Imatge 14. Comparativa connectors PATA i SATA

4. Conclusions

En aquest projecte hem vist com la tasca de dur a terme un projecte de preservació digital no és fàcil. Hi intervenen molts de factors, tant tècnics com humans o institucionals. Això fa que moltes institucions que comencen a desenvolupar un pla de preservació no acabin per implantar-lo degut a les dificultats que els hi representa. Així doncs no fan res amb el seu repositori documental la qual cosa és nefasta a llarg termini. Hem vist la gran quantitat de riscos als que s'enfronta un repositori documental actual i, sense cap actuació, molt probablement les dades s'acabin per perdre definitivament, impactant molt negativament en la viabilitat de la feina de la institució. Què seria d'una biblioteca sense llibres que consultar?

La actual dependència digital de les empreses i institucions fa que no s'hagi d'esperar molt més per començar a implantar un sistema de preservació, i el problema es fa més greu any rere any ja que cada vegada es generen més i més documents digitals, molts d'ells crítics per a la feina de la empresa o institució.

Actualment els problemes als que s'enfronta un projecte de preservació digital no són tècnics ni de personal, són principalment econòmics i institucionals. Moltes institucions afronten la pèrdua de dades amb accions específiques però aïllades unes d'altres (contractació temporal de personal específic, pressuposts extraordinaris, accions tècniques puntuals…) però no compten amb un vertader pla a llarg termini per preservar les seves dades. Aquesta falta de continuïtat i sistematització fa que la problemàtica de la pèrdua de recursos no desaparegui i que tard o d'hora s'acabi manifestant.

Així doncs, consideram clau la implicació institucional i una adequada dotació econòmica, que en molts de casos serà elevada, la qual cosa dificulta la seva implantació total, però que encara amb poc pressupost es pot començar a desenvolupar polítiques i bones pràctiques de preservació. Lo més important és que no siguin pràctiques aïllades i que corresponguin a un pla ben definit a llarg termini amb la implicació de gerència i els responsables de la institució.

Amb la execució d'aquest projecte s'han assolit coneixements de una de les matèries de la informàtica de la qual se'n parla poc. La preservació digital hauria d'estar ja en fase d'implantació a la majoria d'institucions i la realitat diu que no és així. Més enllà de còpies de seguretat o redundància, no hi ha un vertader pla de preservació i això posa en perill la continuïtat de la empresa.

M'hagués agradat haver aprofundit una mica més en les actuals solucions que hi ha al mercat i investigar el seu funcionament a baix nivell, però només he estudiat les seves característiques a alt nivell.

La planificació inicial ha estat més o manco la seguida, tot i que, a mesura que avançava en el projecte he hagut de fer algunes modificacions per dotar de millor coherència al projecte, ja que és quan es desenvolupa tota la informació que s'aprenen coses que inicialment no es veien, com ara la estructuració de la part pràctica, que en un primer moment era molt genèrica i després d'haver estudiat el model de referència OAIS, vaig preferir adaptar-la a aquest model.

5. Glossari

ATA: Advanced Technology Attachment
CD: Compact Disc
DVD: Digital Video Disc
ISO: International Organization for Standardization
GNU: GNU's not UNIX
LOPD: Llei Orgànica de Protecció de Dades
METS: Metadata Encoding and Transmission Standard
OAIS: Open Archival Information System
OCR: Optical Character recognition
PATA: Parallel ATA
PREMIS: PREservation Metadata: Implementation Strategies
SAI: Sistema d'alimentació ininterrompuda
SATA: Serial ATA
UMD: Universial Media Disk
XML: eXtensible Markup Language

6. Bibliografia

- Miquel Térmens Graells, Preservación digital, Editorial UOC
- Alice Keefer, Núria Gallart, La preservación de recursos digitales, Editorial UOC

- La gestión de los recursos digitales en una biblioteca virtual: la Biblioteca virtual de la UOC
  http://openaccess.uoc.edu/webapps/o2/bitstream/10609/317/1/12836.pdf
- Article Wikipèdia:
  https://es.wikipedia.org/wiki/Preservaci%C3%B3n_digital
- http://www.kultura.ejgv.euskadi.eus/contenidos/informacion/kultura2_0_prestakuntza/es_k20_form/adjuntos/cuaderno-DIG-52.pdf
- Procés de digitalització de la Biblioteca Nacional d'Espanya
  http://www.bne.es/webdocs/Catalogos/ProcesoDigitalizacionBNE.pdf
- PRESERVACIÓN DIGITAL DE FICHEROS MÁSTER DE PRENSA ACTUAL
  http://e-archivo.uc3m.es/bitstream/handle/10016/17889/uriarte_preservacion_TFM_2013.pdf?sequence=2
- Gestión documental para gente normal
  https://gestiondocumentalparagentenormal.com/
- Diccionario de Datos PREMIS de Metadatos de Preservación
  http://www.bne.es/es/Micrositios/Publicaciones/PREMIS/index.html
- Software XENA
  http://xena.sourceforge.net/
- Software Rosetta
  http://www.exlibrisgroup.com/category/RosettaOverview
  http://library.harvard.edu/rosetta
- LOPD:
  https://www.agpd.es/portalwebAGPD/canaldocumentacion/informes_juridicos/reglamento_lopd/index-ides-idphp.php
- Llei de propietat intel·lectual
  http://www.mecd.gob.es/cultura-mecd/areas-cultura/propiedadintelectual/la-propiedad-intelectual.html
- Escàners FUJITSU
  http://www.fujitsu.com/global/products/computing/peripheral/scanners/
- Oracle VMWare
  https://www.vmware.com/es/virtualization/overview

- Virtualización:
  http://blog.capacityacademy.com/2012/08/07/que-es-la-virtualizacion-y-cuales-son-sus-beneficios/
- Temperatura discs durs
  http://www.buildcomputers.net/hdd-temperature.html
  https://www.backblaze.com/blog/hard-drive-temperature-does-it-matter/
- SSD
  http://www.ibtimes.co.uk/ssds-lose-data-if-left-without-power-just-7-days-1500402
- Tipus de SAI
  http://computerhoy.com/noticias/hardware/que-es-sai-tipos-sai-cual-elegir-39147
- Causas deterioro discos duros
  http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-77432010000400007
- Estudis d'impacte d'humitat als discs
  http://www.zdnet.com/article/heat-doesnt-kill-hard-drives-heres-what-does/