

Repositoris digitals

Disseny i implementació per a biblioteques, arxius i museus

4/4

Ciro Llueca
Sandra Reoyo

Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya
25 d'octubre de 2012



Programa

1. **Què són els repositoris digitals i per a què serveixen** (Dimarts 16.10.12)
 - 1.1 Cultura *Open Access*: de la *serial crisis* a la resposta de la comunitat
 - 1.2 Tipus de repositoris: institucionals, cooperatius, temàtics, recol·lectors
2. **Disseny i implementació d'un nou repositori** (Dijous 18.10.12)
 - 2.1 Cobertura estratègica i legal: rere la coherència amb la institució i la llei
 - 2.2 Maquinari i programari: sistemes propietaris, lliures i a mida
3. **Difusió i avaluació** (Dimarts 23.10.12)
 - 3.1 Planificació dels processos de treball
 - 3.2 Publicació i promoció: objectiu "repositoris sostenibles"
 - 3.3 Avaluació: visibilitat, interoperabilitat i polítiques públiques
4. **Preservació digital** (Dijous 25.10.12)
 - 4.1 Estratègia: definint una política per al futur
 - 4.2 Tàctica: a l'espera de la solució global, algunes recomanacions

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

1. Què són els repositoris digitals i per a què serveixen (síntesi)

En síntesi (1)

- L'accés obert és un moviment que promou l'accés universal al coneixement finançat amb diners públics
- Va sorgir a Internet com a resposta a una sèrie de circumstàncies econòmiques i culturals que han afectat a les biblioteques, arxius i museus

En síntesi (2)

- Els repositoris digitals són “un conjunto de servicios prestados por una institución a su comunidad para recopilar, gestionar, difundir y preservar su producción digital a través de una colección organizada, de acceso abierto e interoperable”
(Barrueco et al., 2010)

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

En síntesi (3)

- Serveixen per publicar a Internet, normalment amb accés lliure, una base de dades que inclou els documents digitals o digitalitzats que formen els fons i col·leccions relacionades amb una unitat d'informació
- Compartir i preservar!

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

En síntesi (4)

- Existeixen repositoris institucionals, cooperatius, temàtics i portals
- Existeixen passarel·les i sistemes que ofereixen accés a múltiples repositoris

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

2. Disseny i implementació d'un nou repositori (síntesi)

En síntesi (1)

- Les unitats d'informació han de conèixer la planificació estratègica de la institució (i participar, si és possible)
- Iniciar un procés de disseny i implementació d'un repositori digital ha de partir necessàriament d'un pla que sigui coherent amb l'entitat que finançarà i impulsarà la nova eina

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

En síntesi (2)

- Les lleis de propietat intel·lectual vigents en la majoria de països obliguen a obtenir permisos dels propietaris dels drets morals i/o materials per publicar en els repositoris, i en quins termes
- Els administradors de repositoris han de tenir el permís dels autors o propietaris dels drets per difondre les obres incloses

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

En síntesi (3)

- L'elecció entre un sistema local o extern de maquinari depèn de diversos factors, bàsicament econòmics
- Si la institució no té un equip informàtic estable i personal tècnic, serà més ràpid i econòmic començar ubicant el repositori en un sistema extern
- A llarg plaç, però, serà més car mantenir una estructura externa

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

En síntesi (4)

- L'elecció del programari es realitza en base al tipus de document a incloure i el volum previst d'usuaris
- No existeix un únic programari amb solucions per tots els casos
- DSpace, un programari lliure, és el més usat

Llueca & Reoyo. Repositoris digitals. COBDC oct 2012.

3. Difusió i avaluació (síntesi)

En síntesi (1)

- En paral·lel a comptar amb cobertura estratègica i legal, i a definir les necessitats de maquinari i programari, és necessari:
 - Definir les col·leccions que formaran part del repositori (pensant en clau d'impacte)
 - Identificar els recursos que ens calen per a la implementació
 - Determinar el sistema de metadades, essencial per a la seva futura interoperabilitat i visibilitat

En síntesi (2)

- La majoria de repositoris compten amb el protocol d'interoperabilitat creat per la Open Archives Initiative, el Protocol for Metadata Harvesting (OAI-PMH)
- Aquest protocol serveix per capturar (si el repositori és agregador) o proveir (si és proveïdor) metadades de/a altres repositoris

En síntesi (3)

- Les estratègies de comunicació han de considerar als dispositants potencials i també als usuaris
- Establir aliances és imprescindible: directius, departaments paral·lels, unitats d'informació, altres repositoris, etc.
- Ús de la web social (amb un pla d'acció)

En síntesi (4)

- Existeixen diverses guies per avaluar els repositoris
- Són eines a considerar abans, durant, i després de la implementació
- Alguns dels aspectes a avaluar: la visibilitat, accés a les polítiques del repositori, aspectes legals, interoperabilitat, etc.

4. Preservació digital

- 16:00h Síntesi de la classes anteriors
- 16:15h Panorama
El digital: molta informació, poca durabilitat
- 16:45h Estratègia
Definint una política per al futur
- 17:45h Pausa
- 18:15h Tàctica
A l'espera de la solució global, algunes recomanacions

4. Preservació digital. Panorama

Panorama (1)

“Año 2054. Mis nietos (que no han nacido) están en el desván de mi casa (que no he comprado). Descubren una carta de 2007 y un CD-ROM. La carta dice que el CD tiene un documento en el que se da la clave para heredar mi fortuna (que no he ganado). Mis nietos sienten curiosidad, pero nunca han visto un disco compacto, salvo en viejas películas. Aun cuando localizaran un lector de disco adecuado, ¿cómo lograrían hacer funcionar los programas necesarios (como Microsoft Word) para interpretar el contenido?”

“Si no fuera por la carta, mis nietos no sabrían que deben abrir el CD-ROM. La carta es legible, sin máquinas ni conocimientos a parte del propio idioma. Dado el carácter mutable de programas y circuitería, dentro de 50 años la única información legible será la carta.”

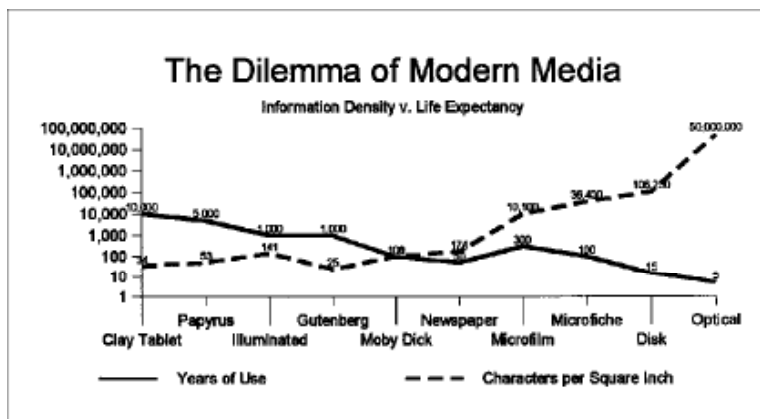
Jeff Rothenberg. “¿Son perdurables los documentos digitales?”. *Investigación y ciencia*, (marzo 1995). <http://www.dfists.ua.es/~gil/articulo.pdf>

(Keefer, Gallart, 2007)

Panorama (2)

Documents primitius: menys informació, més durada.

Documents digitals: més informació, menys duradera.



(Paul Conway. *Preservation in the digital world*, 1996).

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

Panorama (3)

- Anys 80-90: els responsables de preservació tradicional observen amb preocupació l'augment de recursos digitals en els seus fons.
 - Saben com preservar documents analògics, però desconeixen què fer amb els nous **suports** (CD-ROM, DVD, accés en línia) i **formats** (text, imatges, hipertext, multimèdia).

“Disponemos de cintas de 1949 que suenan perfectamente, y otras de 1989 que se han perdido completamente”

(Marc Kirkeby, Sony, en relació a l'accés a 600.000 enregistraments musicals en cintes magnètiques)

(Keefer & Gallart, 2007)

Panorama (4)

- Biblioteques, Arxius i Museus no han estat immunes als canvis tecnològics i la irrupció d'Internet:
 - D'una banda, han d'integrar a les seves col·leccions recursos digitals, en formats i suports que no estan consolidats. Han de fer-ho a gran escala. Són els documents **nascuts digitals** (*born digital*).
 - D'altra, han d'aprofitar les oportunitats del medi digital per difondre les seves col·leccions de **documents analògics**, i contribuir també a la seva preservació.

Integración de los *born digital*



Difusión (y preservación) de analógicos



Movimientos Migratorios



<http://pares.mcu.es/MovimientosMigratorios>

Movimientos Migratorios

- El *portal Movimientos Migratorios Iberoamericanos* coordinat pel **Ministerio de Cultura** amb l'objectiu de facilitar l'accés a la documentació administrativa relativa a la emigració espanyola a Iberoamèrica, i que compta amb la col·laboració de l'**Archivo General de la Nación de México**, l'**Archivo General de la Nación de la República Dominicana**, l'**Archivo General de la Nación de Uruguay**, i la **Fundación Complejo Cultural Parque de España**, d'Argentina, i és prevista la incorporació de l'**Archivo Nacional de la República de Cuba**.
- El sistema inclou registres dipositats directament en el repositori, o bé dipositats en d'altres repositoris associats, sempre que compleixin metadades estàndars.
- Software *a mida*, 34.000 registres.

Panorama (5)

- "Abarca todas las consideraciones relacionadas con la gestión, administración, financiamiento y recursos humanos que tienden a salvaguardar el bienestar de las colecciones de bibliotecas" (IFLA)
- "Actividades asociadas con el mantenimiento en buenas condiciones del material de archivo y bibliotecario para su utilización en forma original" (ALA)
- "La preservació a llarg termini implica la provisió d'accés indefinit als documents o, com a mínim, al contingut d'aquests (...). En el document d'arxiu, la preservació del context i entorn de creació explicarà, complementarà i donarà valor al document" (Soler, 2009)

(Soler, 2009)
(Keefer, Gallart, 2007)

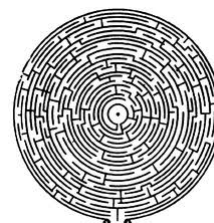
Panorama (6)

- **"Preserving digital information"** (Research Libraries Group & Commission on Preservation and Access, USA, 1996)
- **"Digital preservation needs and requirements in RLG member institutions"** (Hedstrom & Montgomery, 1998)
- **"Preservation in the digital world"** (Yale University, USA, 1996)
- **"Long term preservation of electronic materials"** (Joint Information Systems Committee & British Library, UK, 1996)
- CEDARS Digital Preservation Project (Granger et al., 1998-2002)
- CAMALEON (University of Leeds, UK, Rothenberg)
- **"PADI: Preserving Access to Digital Information"** (National Library of Australia, 1996)
- **"Directrices para la preservación del patrimonio digital"** (UNESCO, 2003)
- **"NEDLIB", "E-Depot" & "DIAS"** (Koninklijke Bibliotheek, DK, 1998-2000)
- **"Kulturarw3"** (Kungliga Biblioteket, SE, 1996)

(Keefer, Gallart, 2007)

Panorama (7)

- Després de diversos estudis i projectes, no existeix en l'actualitat un sistema informàtic que compleixi tots els requeriments de preservació digital
- Es combinen estratègies a mig plaç i activitats experimentals per garantir l'accés als fons digitals



Panorama (8)

- Propietatst de la preservació digital:
 - És una **sèrie d'activitats**, no tècniques aïllades.
 - La finalitat és la **utilització futura** dels documents.
 - Es pretén conservar el **document físic, a més de la informació** que conté.

(Keefer & Gallart, 2007)

4.1 Estratègia

Estratègia (1)

- La preservació d'un fons tradicional s'establia en base al material a conservar i al temps de preservació



Les "Homilies d'Organyà" (finals s. XIII) és un dels escrits més antics en català. Descobert el 1904, es conserva a la Biblioteca de Catalunya

(Keefer, Gallart, 2007)

Estratègia (2)

- A diferència de l'actuació tradicional, orientada a mantenir el document en el seu estat original, els documents digitals han de ser modificats i transformats contínuament per a que continuïn sent accessibles i intel·ligibles
- Es tracta de **preservació per alteració**
 - A més de preservar una versió accessible i llegible, els experts recomanen preservar, a més, la cadena de bits original

(Keefer, Gallart, 2007)

Estratègia (3)

- La preservació digital és més complexe:
 - Quins recursos cal preservar?
 - Quan s'ha d'iniciar el tractament?
 - Qui s'implicarà en el procés de preservació?



<http://www.casadelabiblia.org/>
[consulta octubre 2012]

(Keefer, Gallart, 2007)

Estratègia (4)

Preservació tradicional (Baró, 1990)	Preservació digital (Jones, Beagrie, 2001)
- Obligació legal - Valor intrínsec del material - Pertinença als interessos de la institució - Estat físic - Ús actual i potencial - Valor artístic - Política sobre emplaçament i manipulació - Disponibilitat en altres biblioteques - Cost-Efectivitat del tractament - Disponibilitat del personal i instal·lacions - Existència de suports alternatius	Les anteriors, i definir si: - El material s'ajusta a la política de selecció - Una altra institució ha assumit la responsabilitat - Hi ha obligació legal - El valor del document justifica la inversió - La institució pot manipular el document - El recurs és llegible avui i en el futur - El recurs està en un suport que serveix per transferir o emmagatzemar les dades - El recurs disposa de documentació identificativa, incloent metadades

7)

Estratègia (5)

- Tot i existint capacitat tècnica i voluntat institucional de preservació digital, els recursos limitats obliguen a les institucions a establir una política:
 - Necessitats i recursos de la institució
 - Interessos dels usuaris
 - Diferents aspectes físics i intel·lectuals

(Keefer, Gallart, 2007)

Estratègia (6)

- L'estratègia requereix trobar l'equilibri entre la importància relativa del document i els factors de risc



Disculpa, no encontramos la página solicitada.

Es posible que ya no exista en el sitio, haya cambiado de ubicación o no esté disponible temporalmente.

Intenta lo siguiente:

1. Verifica que hayas escrito correctamente la dirección.
2. Haz clic en la opción **Atrás** para intentar otro vínculo.
3. Utiliza el **mapa de sitio** del portal para ubicar la ruta de la página que deseas consultar.
4. Introduce en el campo de búsqueda una palabra clave que te ayude a encontrar la página que necesitas.

(Keefer, Gallart, 2007)

Estratègia (7)

- Actuació reactiva o proactiva?
 - La solució reactiva no és recomanable per objectes digitals: al detectar l'error, sol ser massa tard
 - El “llarg termini” en objectes digitals és de 20 anys (Waller, Sharpe, 2006)



1986



1987



1996



2006

(Keefer, Gallart, 2007)

Estratègia (8)

- Actuació esporàdica o continuada?
 - La solució esporàdica permet allargar la vida dels documents analògics, però difícilment dels digitals
 - Les intervencions digitals afecten al document i a la seva descripció
 - La intervenció continuada requereix financiació permanent



(Keefer, Gallart, 2007)

Estratègia (9)

- Actuació o omissió?
 - L'abandonament benigne (*benign neglect*) servia en alguns casos per salvar la vida del document
 - En l'entorn digital, l'oblit és letal. Ha de ser llegible el format i també el suport



Sinopsis

Material fotográfico inédito de los insignes fotógrafos Robert Capa, Gerda Taro y David Seymour aparecieron en varias cajas de un apartamento de México D.F. Esta magnífica serie de negativos representan una extraordinaria colección de imágenes que representan diferentes escenas de la contienda civil española (1936-1939) en el frente de batalla, donde se capta perfectamente el horror, el caos de la guerra, de cómo el fascismo y el comunismo se enfrentan en el preludio de la II Guerra Mundial.

WordStar is highly flexible and very visible. Watch the screens as you give commands, and information in various parts of the screen will guide you. You won't see all the information all the time, but it will be there when you need it.

Wordstar 1992

(Keefer, Gallart, 2007)

Estratègia (10)

- Actuació al crear el document o en cas de deteriorament?
 - És recomanable, tot i els reptes (creació de documents descentralitzada, inèrcia dels autors i departaments) establir pautes en la creació, i assessorar (selecció de formats, suports, metadades)



(Keefer, Gallart, 2007)

Estratègia (11)

- Hem de decidir:
 - Preservar el contingut, inclús millorar la llegibilitat o utilitat mitjançant les tecnologies disponibles en el moment de la seva recuperació
 - Mantenir el recurs mitjançant la preservació de l'aspecte i comportament original, amb les limitacions que poden representar



(Keefer, Gallart, 2007)

Estratègia (12)

- Valors essencials a preservar en documents d'arxiu:
 - Fiabilitat: és possible conèixer la traçabilitat, des de que fou creat per algú, fins el present.
 - Autenticitat: deixant clar què és el màster i què còpies de difusió, ús, etc.
 - Integritat: ha de restar intacte, tal com mostren les seves metadades.
 - Identitat: es coneixen totes les persones que han creat o treballat en el document.
 - Capacitat d'ús i accés per a l'ull humà.

(Soler, 2009)

(Elvira, 2011)

Estratègia (13)

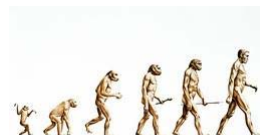
- Prioritzant:
 - Documents nascuts digitals
 - Documents digitalitzats sense versió analògica
 - Documents digitalitzats amb versió analògica

(Keefer, Gallart, 2007)

4.2 Tàctica

Tàctica (1)

- Existeixen dues accions bàsiques de preservació digital, que habitualment actuen combinades:
 - Actualització dels suports
 - Migració dels formats



Tàctica (2)

- Actualització / *Refreshing*
 - Transferència de dades d'un suport a un altre
 - No preserva les dades, però és un pas repetible i imprescindible per garantir l'accés a aquestes
 - Quina previsió de vida té el nou suport?



(Keefer, Gallart, 2007)

Tàctica (3)

- Migració
 - Transferència de dades d'un **format** a un altre
 - L'estratègia més usada, tot i existir el risc de perdre elements del document
 - És una operació experimentada, sense coneixements tècnics especialitzats
 - Es difícil programar la migració en objectes complexos, altera els documents, i s'ha de repetir al llarg del temps



(Keefer, Gallart, 2007)

Tàctica (4)

- Emulació
 - **Simular** el comportament del programari original
 - L'estratègia més respectuosa amb el document original
 - No requereix un seguiment continuat del format, soluciona la lectura per objectes complexos
 - Existeixen poques experiències reals, els emuladors han d'actualitzar-se, requereix coneixements especialitzats ara i en el futur

(Keefer, Gallart, 2007)



Tàctica (5)

- Accions complementàries:
 - **Museu tecnològic:** conservar tots els models, còpies i peces de reparació, per poder emular
 - **Replicació:** preservar múltiples còpies digitals en línia, en formats estàndars (Castillo, Jorba, 2008)
 - **Analogització:** apostar per suports de solvència contrastada en preservació: paper, microformes, etc.
 - **Arqueologia digital:** confiar en el futur

(Keefer, Gallart, 2007; Serra 2009)



Tàctica (6)

- Riscos per la longevitat:
 - Visualització: en el futur hauran de conèixer la infraestructura original
 - Interferències: a causa d'enciptació o compressió, la recuperació pot ser problemàtica
 - Interrelacions: els enllaços d'hipertext es poden perdre
 - Custòdia: qui es responsabilitza?
 - Traducció: salvar el *look & feel*

(Besser, citat per Keefer, Gallart, 2007)



CLAMOR

English version Qué es la Fundación

Fundación Juan March

Inicio Arte Música Conferencias Ciencias Sociales Biblioteca

INICIO > MÚSICA > CLAMOR >

CLAMOR
COLECCIÓN
DIGITAL DE
MÚSICA
ESPAÑOLA

Conciertos

Clamor
Clamor es un portal de conocimiento que aspira a ir incorporando toda la actividad concertística de música española desarrollada en la Fundación Juan March desde su inicio en 1975. Clamor enriquece el audio de las grabaciones en vivo,...

Acceso a la colección

- Conciertos
- Compositores
- Obras

Compositores

Más de 300 compositores españoles en

<http://digital.march.es/clamor/>

CLAMOR

- La Fundació Juan March creada al 1955 pel financer espanyol Juan March Ordinas. És una institució privada que desenvolupa les seves activitats en el camp de la cultura humanística i científica. Té seu a Madrid, y és titular del Museo de Arte Abstracto Español (Cuenca) i del Museo Fundación Juan March (Palma de Mallorca)
- Al 2011 ha obert el seu repositori digital: CLAMOR - Colección Digital de Música Española, amb l'objectiu de recopilar, preservar i donar accés a la activitat concertística de la Fundació, des de 1975
- Conté gravacions en viu, programes de mà, informació sobre els intèrprets, fotografies, etc.
- Fedora (adaptació Islandora)

El suport (1)

- Factors de degradació:
 - La composició “en capes” dels suports, en les quals s’energistren (un cop i un altre) les dades per mitjà d’un procés magnètic poden causar envelliment precoç, que pot provocar errors en la lectura de les dades.

(Keefer, 2012)

El suport (2)

- Factors de degradació:
 - Temperatura i humitat
 - Baixa qualitat dels materials originals
 - Baixa qualitat del procés de fabricació
 - Falta de controls ambientals en la fabricació, transport i emmagatzegament
 - Proximitat a camps magnètics
 - Brutícia, pols, en l’entorn
 - Introducció de bacteris, fongs
 - Mal estat dels equips d’enregistrament i de lectura

(Keefer, 2012)

El suport (3)

- Alguns factors de degradació són controlables per la institució:
 - Flexibilitzant els criteris de compra, que dificulten l'adquisició de material d'alta qualitat
 - Promovent el canvi en la inèrcia o reticència de les maneres de treballar
 - Facilitant la capacitat tècnica o econòmica per efectuar canvis en l'entorn físic

(Keefer, 2012)

El suport (4)

Esperança de vida segons condicions ambientals (humitat relativa i temperatura):

Suport	25 h.r. / 10 C	40 h.r. / 20 C	50 h.r. / 28 C
D3 cinta magnètica	50 anys	15 anys	1 anys
DLT cinta magnètica	75 anys	15 anys	1 anys
CD/DVD	75 anys	20 anys	2 anys
CD-ROM	30 anys	3 anys	3 anys

(Síntesi de Jones & Beagrie, 2001, p. 130)

El format (1)

- L'elecció d'un format pot condicionar:
 - La selecció o no del recurs per ser preservat.
 - El mètode usat per la seva preservació.
 - La facilitat de recuperar i visualitzar el document.
 - El tipus d'ús possible en el futur.
 - El cost de la preservació.

(Keefer, 2012)

El format (2)

- La importància d'escollir un format:
 - Format estàndard, reconeixible per una àmplia gama de programes (.txt)
 - Format propietari amb gran base de clients, que puguin arribar a incentivar econòmicament les solucions (.doc)
 - Format propietari de difusió limitada (.dxf)

(Keefer, Gallart, 2007)

El format (3)

- Existeixen ca. 16.000 formats (NARA)

- Document (doc, rtf, pdf, xsd, ps)
- Imatge (jpeg, gif, png, psd, tif)
- Audio (wav, mp3, aac)
- Video (wmv, avi, rm, mpg)
- Full de càlcul (xls, xsc)
- Estadística (SPSS)
- Diagrama o disseny (Visio, AutoCAD)
- Bases de dades (SQL, MySQL, Oracle, Access)
- Presentacions (ppt)
- Pàgines web
- Etc.



<http://www.gdfr.info/>

The technical registry
PRONOM

(Bates et al., 2006, citat per Keefer, Gallart, 2007)

El format (4)

Formats	Preferit	Acceptable	Problemàtic
Dades estadístiques	SPSS, por, sav	STATA, xport, text separat per tabulació	Text d'amplada fixe (no delimitat)
Base de dades	SQL separat per tabuladors o comes	ACCESS, dBase, FilemakerPro, FoxPro, Paradox	Text d'amplada fixe (no delimitat)
Full de càlcul	Text delimitat separat per tabuladors	Excel, Lotus, Quattro Pro	
Text	Enrquit, rtf, xml, txt	MsWord, HTML, Wordperfect, paper	CAQDAS, WinMax, Nudist
Imatge	tiff, png	gif, bmp	Adobe PDF
Video	MPEG	Quicktime	
So	mp3	wav	

(Recomanacions UK data Archive, citat per Keefer, Gallart, 2007)

Preservem digitalitzant?(1)

- La digitalització s'usa com a mètode de preservació (i difusió) dels documents analògics, atès que pot promoure:
 - Cerca i opcions de localització més fàcils.
 - Accés sense barreres temporals, físiques o geogràfiques.
 - Més maneres de visualitzar el document.
 - Millor qualitat de la imatge i el text.
 - Fàcil i ràpida reproducció, sense agressions a l'original.
 - Nous usos, com l'anàlisi d'imatge i text, comparació de documents allunyats, etc.

(Keefer, 2012)

Preservem digitalitzant?(2)

- Pero lògicament, la digitalització no substitueix als processos preservació en els documents analògics:
 - Baixa manipulació.
 - Temperatura i humitat constants.
 - Condicions lumíniques.
 - Insectes, desastres naturals, robatoris o pèrdues, etc.

(Keefer, 2012)



Preservem digitalitzant?(3)

- Prioritzemos?
 - Documents fràgils
 - Singularitat
 - Interés per a la recerca
 - Requeriments d'ús
 - Completar col·leccions
 - Freqüència d'ús i de consulta
 - Domini públic

(Serra, 2011)



Preservem digitalitzant?(4)

- Preparar l'escàner i els documents:
 - Neteja de les lents i de l'entorn de l'escàner
 - Neteja i ordre dels documents.
- Ajustar els aspectes tècnics:
 - Blanc i negre: manuscrits, documents impresos.
 - Escala de grisos: manuscrits, música impresa.
 - Color: mapes, fotos, il·lustracions.



www.digitalpreservation.gov

Preservem digitalitzant?(5)

Tipus document	Preservació	Accés
Text imprès BN/grisos (llibres, revistes)	TIFF 300 dpi	JPEG, PDF 150 dpi
Text imprès color (manuscrits, llibres il·lustrats)	TIFF 300 dpi	JPEG, PDF 150 dpi
Fotografies, mapes	TIFF o JPEG2000 500 dpi	JPEG, PDF 300 dpi
Sonors	WAVE o BWF	MP3
Audiovisuals	MPEG2 o AVI	MPEG2

Més concrecions a:

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

Preservem digitalitzant?(6)

- Accés i emmagatzematge:
 - Carpetes ben organitzades, estructura arbòria, noms i dates clares.
 - Un mínim de 2 còpies en dos suports diferents, de qualitat fiable, allunyats entre si.
 - Etiquetatge curós (mínims: impressió de pantalla per obtenir un sumari, que acompanya les còpies externes al PC.)
 - Emmagatzematge segur.
 - Noves còpies cada 5 anys, com a mínim



www.digitalpreservation.gov

L'arxiu de la Web



Metadades (1)

- Sense metadades, el document “preservat” digitalment pot ser il·legible.
 - Metadades generades automàticament.
 - Metadades generades pels creadors.
 - Metadades generats pel personal de la biblioteca, arxiu, etc.



Metadades (2)

- Sense les metadades no podem garantir que un recurs sigui recuperable i usable:
 - Sense dades descriptives, no es podrà localitzar.
 - Sense dades tècniques, no es podrà visualitzar o utilitzar.
 - Sense dades administratives, no es tindrà la garantia de la integritat ni autenticitat.
 - Sense dades legals, els responsables de la custòdia no sabran quins usos són admissibles o no.

Metadades (3)

- Dades sobre dades
- Metadada és aquella informació descriptiva sobre el context, qualitat, condició o característica d'un recurs documental, dada o objecte, amb l'objectiu de facilitar la seva recuperació, autenticació, avaluació, preservació i/o interoperabilitat

(Wikipedia)

Metadades (4)

- Els documents que s'introdueixen en el repositori han d'estar descrits usant un conjunt de metadades basades en estàndards internacionals per a facilitar el seu ús posterior
- Han d'estar mínimament normalitzades
- Són, en part, l'equivalent a la descripció bibliogràfica en un catàleg clàssic



Metadades (5)

- Tipus de metadades
 - Descriptives: creador, títol, data de publicació, tipus de document, etc.
 - Administratives: propietat, drets i usos, equivalent analògic, etc.
 - Tècniques i de preservació: format, resolució, compressió, dispositiu de captura, estructura del document, historial de canvis, etc.
 - Estructurals: relacions amb altres objectes



Metadades (6)

- Són sistemes estàndards de metadades:
 - DC (Dublin Core)
 - EAD (Encoded Archival Description) > arxius
 - MARC 21 > biblioteques / catàlegs
 - METS (Metadata Encoding and Transmission Standard) > tècniques, preservació
 - PREMIS (Preservation Metadata: Implementation Strategies) > preservació, traçabilitat
 - MODS (Metadata Object Description Schema)
 - TEI (Text Encoding Initiative)

PREMIS

MODS



<ead>



Metadades (7)

- El **Dublin Core** (1995) és l'estàndard més usat, especialment la seva versió **qualificada**
- Pràcticament tots els programaris de repositoris l'inclouen o el suporten
- Consta de 15 elements descriptius:
 - Títol, creador, matèries i paraules clau, descripció, editor, contribucions, data, tipus de recurs, format, identificador, font, llengua, cobertura, gestió dels drets



Metadades (6)

- L'elecció del sistema de metadades és essencial per a la seva futura interoperabilitat i visibilitat, més enllà de la nostra pròpia institució
- El sistema d'autoarxiu pot tenir conseqüències en la correcta assignació de metadades
- Habitualment en l'autoarxiu s'opta per uns mínims (autor, títol, data, etc.), que després completaran els serveis tècnics



Metadades (7)

- La majoria de repositoris compten amb el protocol d'interoperabilitat creat per la Open Archives Initiative, el Protocol for Metadata Harvesting (OAI-PMH)
- El protocol serveix per capturar (si el repositori és agregador) o proveir (si és proveïdor) metadades de/a altres repositoris

<http://www.openarchives.org/pmh/>



Metadades (8)

- Si en el repositori es barregen diferents tipus de col·leccions (científica, docent, cultural, institucional, etc.) és aconsellable diferenciar-les:
 - Per intercanvis futurs
 - Per participar amb cada col·lecció en diferents repositoris



Metadades (9)

- Per participar en portals/agregadors és necessari complir determinades especificacions
 - A DRIVER només documents de recerca
 - A OpenAIRE només documents finançats publicament pel 7PM
- Aquests portals/agregadors es limiten a la integració i recuperació de metadades: el text complet sempre està en el repositori d'origen (cobertura legal)





- El CSIC (Centro Superior de Investigaciones Científicas) és la major agència pública espanyola d'investigació científica (3a UE), amb 15.000 persones en plantilla (d'aquestes 3.000 són investigadors), genera el 20% de la producció científica espanyola
- Va signar la Declaració de Berlín al 2006
- El repositori Digital CSIC (2008-) facilita l'accés a articles (especialment *postprint*), llibres, literatura gris, patents, bases de dades, comunicacions a congressos, material didàctic i divulgatiu, informes tècnics, etc.
- DSpace, 38.000 registres, 180.000 visites mensuals, #2 ranking repositorios (.es)
- Els incentius per als investigadors: augmentar la visibilitat de la seva producció científica; preservar tota la producció; tractament centralitzat (40% oficina tècnica + 40% biblioteques + 15-20% autoarxiu)

Repositori segur (1)

- La majoria de repositoris institucionals no poden ser considerats com repositoris de preservació
- Un repositori de preservació té com a responsabilitat única dur a terme la preservació a llarg plaç dels seus objectes digitals (PREMIS 2008)

(Termens, 2009)

Repositori segur (2)

- Establint un repositori segur:
 - Suports:
 - Emmagatzemar els màsters en suports digitals d'alta qualitat
 - Comprovar la legibilitat dels suports periòdicament
 - Còpies de seguretat:
 - Copiar els màsters i dipositar en lloc extern segur
 - Dispositius d'enregistrament i de lectura
 - Seguiment per garantir que siguin de bona qualitat, i estiguin en bon estat

(Establishing a central depository..., p. 16-17

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

Repositori segur (3)

- Establir un repositori segur:
 - Emmagatzegament:
 - Conservar en entorn controlat de temperatura i humitat
 - Mantenir nivells constants més que seguiment estricte de nivells òptims
 - Allunyat de camps magnètics
 - Mantenir net l'entorn d'operacions
 - Minimitzar la manipulació i l'ús dels suports
 - Seguretat:
 - Control d'accés
 - Accés tancat
 - Protegir l'autenticitat dels fons: no trencar la cadena de custòdia

(Establishing a central depository..., p. 16-17

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

Repositori segur (4)

- Establir un repositori segur:
 - Refresh:
 - Passe periòdic a nous suports (en comprar nous dispositius d'emmagatzegament i lectura?)
 - Seguir un procés de verificació (checksum) per garantir autenticitat i integritat posterior al refresh
 - Documentació
 - Explicar per escrit tota acció duta a terme durant el procés de manteniment o refresh

(Establishing a central depository..., p. 16-17

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

Repositori segur (5)

- El model OAIS (Open Archival Information System) va ser desenvolupat per la NASA, i es va convertir en la norma ISO 14721:2012
- Serveix per referir-se a conceptes i processos en base a una norma compartida
- No ofereix un disseny específic d'arquitectura, és una fulla de ruta



“Vehículo de cuatro ruedas, capaz de transportar entre dos y ocho personas con una velocidad posible de 130 km/h y un espacio para equipaje y con prestaciones para que el viaje sea más agradable (radio, aire climatizado, asientos cómodos...)” (Gladney 2006)

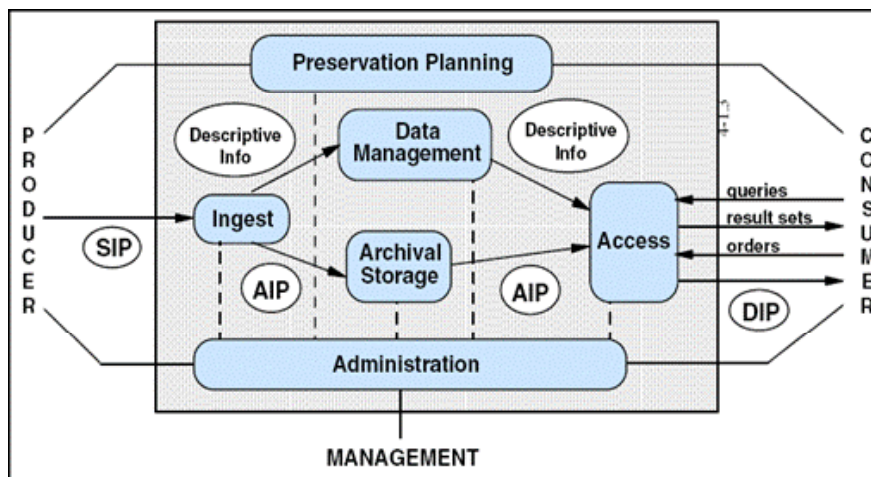
(Keefer, Gallart, 2007)

Repositori segur (6)

- El model OAIS esbossa una arquitectura per al paquet d'informació (objecte+metadades):
 - Ingest: incorporació al repositori
 - Archival storage: emmagatzemament a llarg plaç
 - Data management: la seva gestió
 - Access/dissemination: el seu accés i difusió
 - Preservation planning: preservació planificada
 - Archive administration: gestió del repositori

(Keefer, Gallart, 2007)

Repositori segur (7)



Repositori segur (8)

- Les institucions han de poder ser fiables:
 - Garantitzar la integritat i autenticitat del recurs
 - Vetllar per la protecció dels drets
 - Limitar l'accés a usuaris autoritzats
 - Mantenir les eines necessàries per identificar i recuperar la informació
 - Assegurar la presència de metadades normalitzades

"Trusted digital repositories: attributes and responsibilities" (RLG & OCLC, 2002)

(Keefer, Gallart, 2007)

Repositori segur (9)

- Els repositoris es poden certificar:
 - Per exemple, en planificació i estratègies de preservació, el repositori haurà de:
 - Documentar la seva estratègia de preservació
 - Realitzar estratègies per emmagatzemar i migrar
 - Usar informació tècnica disponible internacionalment
 - Tenir un mecanisme per detectar quan un format està en risc d'obsolescència
 - Etc.

“Audit checklist for the certification of trusted digital repositories” (RLG & NARA, 2005)

(Keefer, Gallart, 2007)

Repositori segur (10)

- Els repositoris es poden certificar:
 - Organització
 - Gestió de l'objecte
 - Infraestructura i seguretat

“Catalogue of criteria for trusted digital repositories” (NESTOR, Network of Expertise in Long-Term Storage of Digital Resources, 2006)

(Keefer, Gallart, 2007)



Repositori segur (11)

- El repositori serà més segur:
 - Si disposa d'un sistema de rèplica en altres institucions
 - Si disposa d'una caixa forta o arxiu fosc, d'accés restringit



Repositori segur (12)

- Externalització dels serveis de preservació:
 - El col·lectiu bibliotecari accepta com un fet impossible que una sola institució pugui afrontar totes les exigències de la preservació digital
 - Més enllà de confiar amb la fiabilitat a llarg plaç dels serveis comercials, s'ha fomentat la creació de centres sense ànim de lucre:
 - Digital Preservation, California Digital Library
 - European Archive
 - UK Data Archive
 - **Metarchive.org**

(Keefer, Gallart, 2007)



CBUC

- Al 2007 va encarregar un estudi a A. Keefer i N. Gallart per garantir la preservació dels documents dipositats en els seus repositoris
- L'informe (en català està disponible a RECERCAT) finalitza amb una sèrie de recomanacions pel CBUC i les institucions participants:
 1. Determinar per a cada repositori quin nivell de preservació es vol garantir
 2. Acordar i documentar les pràctiques sobre obtenció d'autoritats
 3. Acordar i documentar les pràctiques per comprovar la integritat de les obres
 4. Establir recomanacions per crear paquets (objectes + metadades administratives i tècniques)
 5. Decidir un calendari per implementar l'emmagatzematge d'arxiu
 6. Col·laborar amb un mínim de 2 institucions per disposar d'una xarxa de rèpliques anti-desastre, i d'un arxiu fosc
 7. Sol·licitar al soci tecnològic que se sotmeti a certificació i que apliqui la normativa
 8. El·laborar i publicar la política de preservació
 9. Cooperar amb alguna xarxa internacional de prestació de serveis per garantir la preservació
 10. Definir una estratègia de vigilància tecnològica per controlar l'evolució de la tecnologia



CBUC

- El Consorci de Biblioteques Universitàries de Catalunya (CBUC) va ser creat al 1996 com organisme públic autònom per millorar els serveis bibliotecaris a través de la cooperació. Els seus membres són 14 universitats catalanes i quatre organismes d'investigació
- Entre els seus projectes: CCUC, TDX (el primer repositori a Espanya) i d'altres repositoris patrimonials i de producció científica



Lluís Anglada és el seu director: "El naixement del CBUC va ser fruit de diverses coses, de la oportunitat de l'amistat, del moment històric, però sobretot és fill de la voluntat de fer (...) de tenir present que per millorar no n'hi ha prou amb millorar només a casa nostra, sinó que cal millorar tot l'edifici en el que vivim i si podem també el carrer"



MetaArchive: què és?

Una xarxa d'institucions que tenen interès en preservar d'una manera efectiva i amb baix cost alguns dels seus conjunts de dades.

Per fer-ho proposen fer còpies de les dades creuades entre les seves institucions.



MetaArchive: Qui en forma part?

Auburn University
 Boston College
 Clemson University
 Florida State University
 Folger Shakespeare Library
 Georgia Tech
 Indiana State University
 Library of Congress
 Penn State University
 Pontifícia Universidade Católica do Rio de Janeiro (Brasil)
 Rice University
 University of Hull (Anglaterra)
 University of Louisville
 University of North Texas
 University of South Carolina
 Virginia Tech



MetaArchive: Funcionament tècnic

- Usa el programari LOCKSS (Lots of Copies Keep Stuff Safe). Programari molt utilitzat des de fa anys i que ha demostrat la seva efectivitat.
- En un servidor nosaltres guardaríem (no en obert) tesis d'altres universitats i ells les nostres.
- En cas de **pèrdua de fitxers** el sistema les recupera de la xarxa.
- En cas de **corrupció de fitxers** el sistema també els recupera de la xarxa (checksums).



En síntesi (1)

- Després de diversos estudis i projectes, no existeix en l'actualitat un sistema informàtic que compleixi tots els requeriments de preservació digital
- Es combinen estratègies a mig plaç i activitats experimentals per garantir l'accés als fons digitals

En síntesi (2)

- A diferència de l'actuació tradicional, orientada a mantenir el document en el seu estat original, els documents digitals han de ser modificats i transformats contínuament al llarg del temps per a que continuïn accessibles i intel·ligibles
- Es tracta de la preservació per alteració
 - A més de preservar una versió accessible i llegible, els experts recomanen preservar, a més, la cadena de bits original

(Keefer, Gallart, 2007)

En síntesi (3)

- Existeixen dues accions bàsiques de preservació digital, que habitualment actuen combinades:
 - Actualització dels suports
 - Migració dels formats

En síntesi (4)

- El model OAIS esbossa una arquitectura per al paquet d'informació (objecte+metadades):
 - Ingest: incorporació al repositori
 - Archival storage: emmagatzemament a llarg plaç
 - Data management: la seva gestió
 - Access/dissemination: el seu accés i difusió
 - Preservation planning: preservació planificada
 - Archive administration: gestió del repositori

(Keefer, Gallart, 2007)

Lectura recomanada

- Miquel Térmens (Universidad de Barcelona). "Investigación y desarrollo en preservación digital: un balance internacional". *El profesional de la infomación*, vol. 18, núm. 6 (Nov-Dic 2009). <http://eprints.rclis.org/bitstream/10760/16218/1/Investigacion-preservacion-EPI.pdf>

Gràcies!
Estem a la vostra disposició!

Ciro Lluca, clluca@gmail.com
Sandra Reoyo, sandra.reoyo.tudo@gmail.com

