

Universitat Pompeu Fabra
Departament de Comunicació
Taula de Nova Recerca

Digitalització del patrimoni hemerogràfic i la seva preservació

Miquel Térmens Graells
Universitat de Barcelona
Departament de Biblioteconomia i Documentació

termens@ub.edu

Barcelona, 15 de febrer de 2012

1

Programa

1. La conservació de la premsa històrica
2. Tècniques i estàndards de digitalització
3. La digitalització de la premsa històrica
4. La preservació digital
 - 4.1. Preservació de premsa digitalitzada
 - 4.2. Preservació de revistes digitals
 - 4.3. Preservació de diaris electrònics
5. Conclusions

2

1. La conservació de la premsa històrica

- Limitacions per la conservació:
 - Mala qualitat del paper
 - Paper àcid
 - Corrosió per les tintes
 - Condicions ambientals de temperatura i humitat
 - Manipulació humana
- Recomanacions:
 - Limitar la consulta dels originals
 - Crear còpies derivades per a la consulta: microfilms i còpies digitals

3

2. Tècniques i estàndards de digitalització

- Estàndards de digitalització desenvolupats a mitjans de 1990 per:
 - Image Permanence Institute –IPI- (part del Rochester Institute of Technology)
 - Kodak
 - Cornell University
- Punts claus:
 - Òptica
 - Tècniques de compressió d'imatges
 - Capacitat de processament
 - Capacitat d'emmagatzematge
 - Capacitat de transmissió

4

3. La digitalització de la premsa històrica (1)

- Resolució i colors
- El límits de l'OCR (Optical Character Recognition)
- L'estàndard METS-ALTO
- Formats de consulta

- Quin ús li volem donar?
 - Còpia facsímil
 - Còpia optimitzada per a la lectura

STRAUSS Á BARCELONA
Al morir Wagner al any 83, degué anarse'n de aquest mon ab la convicció de que no deixava successor capàs d'empunyar aquella batuta qu'era en las sevas mans lo ceptre de la música moderna.

STRAUSS Á BARCELONA
Al morir Wagner al any 83, degué anarse'n de aquest mon ab la convicció de que no deixava successor capàs d'empunyar aquella batuta qu'era en las sevas mans lo ceptre de la música moderna.

STRAUSS Á BARCELONA
Al morir Wagner al any 83, degué anarse'n de aquest mon ab la convicció de que no deixava successor capàs d'empunyar aquella batuta qu'era en las sevas mans lo ceptre de la música moderna.

STRAUSS Á BARCELONA
Al morir Wagner al any 83, degué anarse'n de aquest mon ab la convicció de que no deixava successor capàs d'empunyar aquella batuta qu'era en las sevas mans lo ceptre de la música moderna.

STRAUSS Á BARCELONA
Al morir Wagner al any 83, degué anarse'n de aquest mon ab la convicció de que no deixava successor capàs d'empunyar aquella batuta qu'era en las sevas mans lo ceptre de la música moderna.

STRAUSS Á BARCELONA
Al morir Wagner al any 83, degué anarse'n de aquest mon ab la convicció de que no deixava successor capàs d'empunyar aquella batuta qu'era en las sevas mans lo ceptre de la música moderna.

Resolució i colors



Sortida de la digitalització

Sortida directa de la digitalització



Pàgines separades



Pàgines amb marca d'aigua



Un fitxer PDF



Gir de pàgines



7

3. La digitalització de la premsa històrica (2)

- Alguns països controlen els programes de digitalització. Exemple els EUA:
 - *National Digital Newspaper Program*. <http://www.loc.gov/ndnp/>
 - *Chronicling America; Historic American Newspapers*. <http://www.loc.gov/chroniclingamerica/>
 - 140.115 publicacions (1690 – present)
 - 4,7 milions de pàgines digitalitzades
- El major projecte de digitalització al Regne Unit:
 - *The British Newspaper Archive*. <http://www.britishnewspaperarchive.co.uk>
 - Consultable: 200 publicacions; 4 milions de pàgines digitalitzades
 - Previst: 40 milions de pàgines (any 2020)

8

3. La digitalització de la premsa històrica (3)

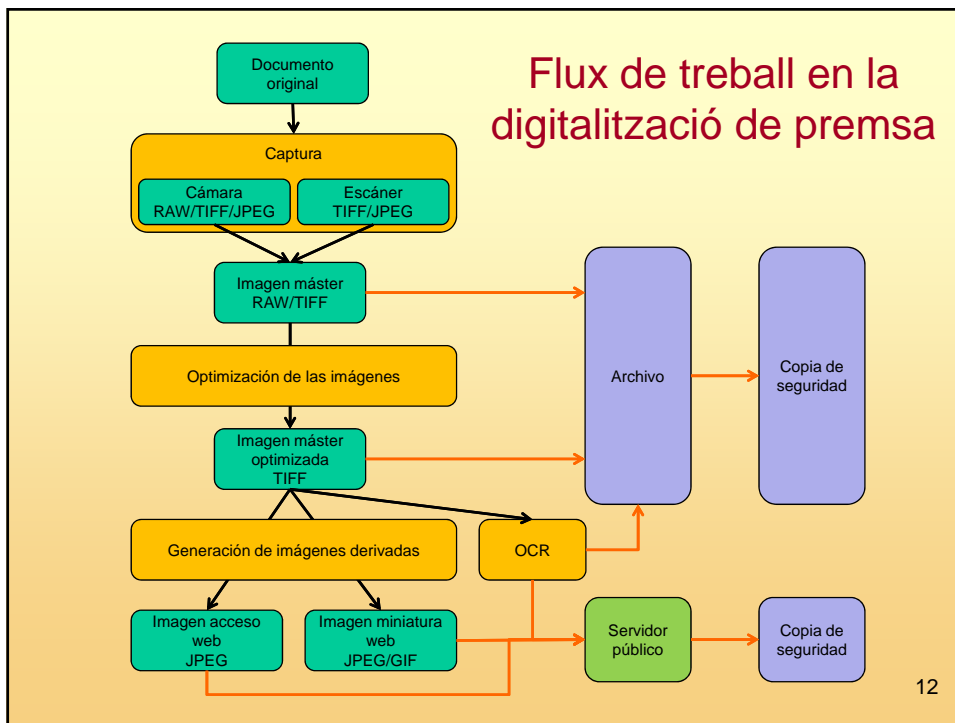
- Models de consulta:
 - Serveis privats: Proquest, JSTOR
 - Dipòsits públics: ARCA
 - Directoris o agregadors: Biblioteca Virtual de Premsa Històrica
- Conflicte d'interessos de les biblioteques amb el sector editorial: volen reservar-se una futura font d'ingressos.

9

Comparació dels sistemes de digitalització



10

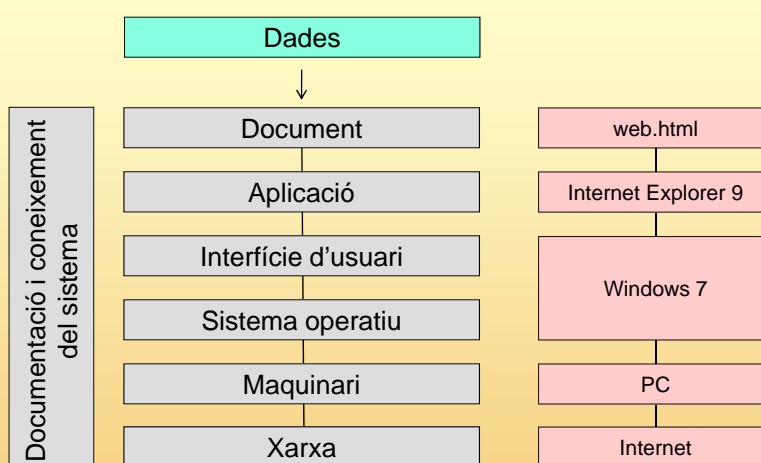


4. La preservació digital

- Recerca i desenvolupament des del 1994.
- Model funcional: OAIS (Norma ISO 14721:2003)
- Formats oberts: XML
- Estàndards de metadades: METS, PREMIS
- Emfasis en:
 - Automatització de les operacions d'ingesta
 - Escalabilitat dels sistemes
 - Sostenibilitat econòmica: reducció de costos i model de finançament

13

La dependència tècnica




14

4.1. Preservació de premsa digitalitzada

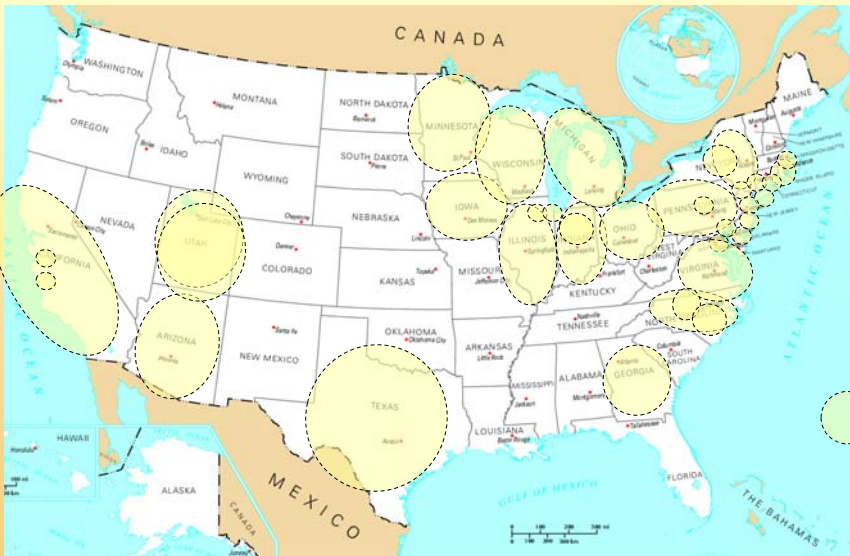
- A càrrec de les biblioteques: directament o en cooperació
- Alternatives:
 - Model xarxa peer to peer: LOCKSS (Stanford Univ.)
 - Model centre de preservació cooperatiu: HathiTrust (Michigan Univ.). 62 institucions



15

 **HathiTrust** <http://www.hathitrust.org/>




Biblioteques digitals de 52 universitats



The map displays the United States with 52 yellow circles of varying sizes indicating the locations of HathiTrust member universities. The circles are distributed across various states, including Washington, Oregon, Nevada, Utah, Arizona, Colorado, New Mexico, Texas, Oklahoma, Missouri, Illinois, Indiana, Ohio, Pennsylvania, Virginia, North Carolina, South Carolina, Georgia, Florida, and Maine. The map also shows neighboring countries Canada and Mexico, and the Atlantic Ocean to the east.

16

4.2. Preservació de revistes digitals

- A càrrec de les editorials, directament o contractada amb tercers
- Alternatives.
 - Model preservació interna: Elsevier 
 - Inversió de 300 M\$ a Ohio (EUA):
 - Datacenter dins d'un búnquer soterrani antinuclear
 - 2 Petabytes d'emmagatzemament
 - 6 motors dièsel amb autonomia de 10 dies
 - Model externalització:
 - Privat: Portico (New York). 136 editors 
 - Públic: e-Depot (Amsterdam). 14 editors 

17

4.3. Preservació de diaris electrònics

- El concepte de diari electrònic:
 - Versió facsímil del paper: PDF
 - Versió web HTML. Ex: Internet Archive
 - Versió mòbil: smartphone (Android...), tauleta (iPad...), ebook (Kindle...)
- Els diaris són empreses privades sota un model comercial:
 - L'arxiu és una font d'ingressos
 - Fins quan? conflicte entre: explotació comercial / propietat intel·lectual / interès públic
 - Quan i com es fa el traspàs a la preservació pública?

18

Quina versió?

19

5. Conclusions

- Ja es disposa de bona part de la tecnologia de preservació.
- Els nous diaris electrònics presenten problemes de caracterització i normalització tècnica.
- No estan resolts els models de sostenibilitat.
- No s'ha resolt l'encaix entre interessos privats i públics.