



Universitat de Barcelona
Facultat de Biblioteconomia i Documentació

Seminari amb la professió
Preservació digital

Aspectes tècnics de la preservació digital

Miquel Térmens Graells
termens@ub.edu

Barcelona, 7 de maig de 2007

1

Sumari

1. La magnitud del problema
2. Problemes a nivell tècnic
3. Què s'està fent?
4. Qui ho està fent?
5. Conclusions

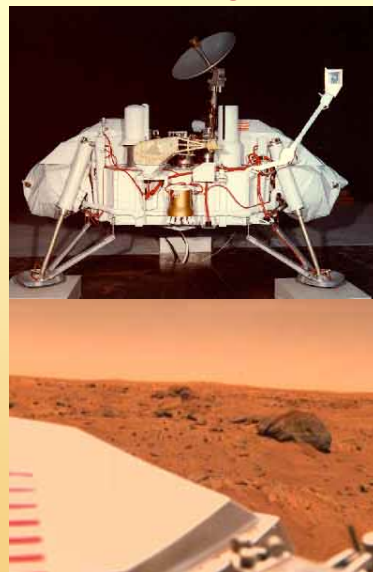
Sumari

1. La magnitud del problema
2. Problemes a nivell tècnic
3. Què s'està fent?
4. Qui ho està fent?
5. Conclusions



Un exemple: NASA Viking

- 1975: la NASA envia les sondes Viking 1 i Viking 2 per analitzar la superfície de Mart.
- 1999: el professor Joseph Miller, de la University of Southern California, vol investigar de nou les dades.
 - Les cintes magnètiques originals són llegibles.
 - Els formats no s'entenen.
 - S'han de teclejar de nou les dades a partir de versions impreses.
- 2001: el reestudi de les dades permet trobar "traces" de vida.



Font: <http://spaceflightnow.com/news/n0107/29marslife/>

La magnitud del problema: xifres

- Google (la web pública) ocupa uns 6 Terabytes.
6.000.000 MB
- Elsevier va digitalitzar el 2001-2004 la totalitat de les seves revistes (des de 1823) = 6 milions d'articles = 50 milions de pàgines = 7 Terabytes.
7.000.000 MB
- La BBC acumula 20.000 hores de video per any que pensen mantenir en digital. Això ocupa uns 2 Petabytes, en formats d'alta qualitat.
2.000.000.000 MB

Què vol dir preservar per sempre?

The Challenge NARA Faces in Preserving Electronic Records:

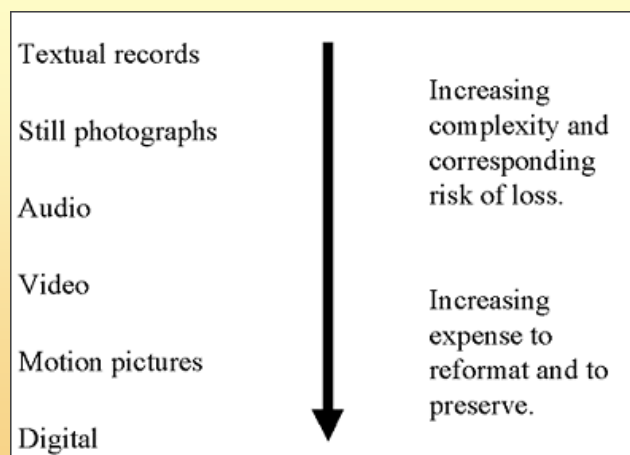
- Preserve any type of electronic record,
 - ✓ Created using any type of application,
 - ✓ On any computing platform
 - ✓ From any entity in the Federal Government, and any donor
 - ✓ Provide discovery and delivery to anyone with an interest and legal right of access
 - ✓ Now and for the life of the Republic

<http://www.archives.gov/era/pdf/thibodeau-planet-storage-2005.pdf>

Sumari

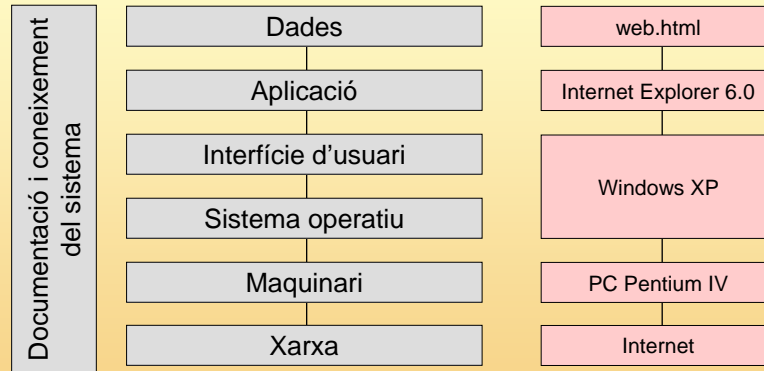
1. La magnitud del problema
2. Problemes a nivell tècnic
3. Què s'està fent?
4. Qui ho està fent?
5. Conclusions

Els mitjans i la preservació



<http://www.archives.gov/preservation/conferences/papers-2003/puglia.html>

La preservació integral, a nivell tècnic



Sumari

1. La magnitud del problema
2. Problemes a nivell tècnic
3. Què s'està fent?
4. Qui ho està fent?
5. Conclusions

La cadena d'estàndards

- Model teòric: OAIS
- Llenguatge d'estructuració: XML
 - Intercanvi de continguts: NLM DTD
 - Codificació descriptiva, administrativa i estructural de metadades: METS
 - Descripció de continguts: MODS
 - MARC21
 - Dublin Core
- Ingesta de fitxers
 - Reconeixement de formats: GDFR
 - Validació de formats: JHOVER
- Formats de preservació: TIFF, PDF-A, AIFF...

Miquel Térmens: Aspectes tècnics de la preservació digital

11

Projectes de preservació

Projectes	Biblioteca nacional	Universitats	Administració	Iniciativa privada
Koninklijke Bibliotheek e-Depot (De Nationale Bibliotheek van Nederland in Den Haag)				
National Archives and Records Administration (NARA). Electronic Records Archives (ERA)				
LOCKSS				
Portico				
Library of Congress (National Digital Information and Infrastructure Preservation Program)				
Internet Archive				
Biblioteca Nacional d'Austràlia				
US Government Printing Office (GPO)				

Miquel Térmens: Aspectes tècnics de la preservació digital


12

Sumari

1. La magnitud del problema
2. Problemes a nivell tècnic
3. Què s'està fent?
4. Qui ho està fent?
5. Conclusions

Qui està fent R+D en preservació?

- NASA
- Library of Congress
- Harvard University
- Stanford University
- Nationale Bibliotheek van Nederland
- Internet Archive



“Multiple approach”: la preservació per Elsevier

- Interna
- Koninklijke Bibliotheek e-Depot (agost 2002-)
- Portico (gener 2006-)
- CLOCKSS (en estudi)

Miquel Térmens: Aspectes tècnics de la preservació digital 16

Preservació interna a Elsevier

“The hardware is a hosting system in Dayton, located in a bunker that is tornado, earthquake and aircraft impact-proof. Daily incremental backups, and weekly complete backups are taken. Off-site copies of backups, and extensive recovery procedures are in place. Migration to a new type of hardware format takes place on every new version release.” (http://www.icsti.org/forum/46/Joep_Verheggen.ppt)

“[Elsevier] has also spent some \$300 million on the company's engine room, a kind of military command center in a bunker in Dayton, Ohio, that has 2 petabytes of data storage (50,000 times the storage of the average household Dell PC) to ensure that a million customers a day get their responses within four seconds. The back-up power supply includes six diesel engine generators with ten days of fuel.” (http://www.forbes.com/global/2002/1111/044_print.html)

Sumari

1. La magnitud del problema
2. Problemes a nivell tècnic
3. Què s'està fent?
4. Qui ho està fent?
5. Conclusions

Conclusions

- El sector privat s'està movent.
- El sector públic (algun) s'està movent.
- La preservació digital serà un (gran) negoci.
- Les solucions tècniques estan arribant. El model OAIS i l'estàndar METS són bàsics.

Preservació...
I vosaltres que esteu fent?

Miquel Térmens Graells
termens@ub.edu