

RISORSE DIGITALI PER LE SCIENZE UMANE  
Università degli Studi di Padova, 8-9 ottobre 2009

# NOTE SULLA CONSERVAZIONE DIGITALE



Maurizio Messina



# Da dove partire?

- **PADI** (*Preserving Access to Digital Information*)  
<http://www.nla.gov.au/padi/>



- **DigitalPreservationEurope**  
<http://www.digitalpreservationeurope.eu/>



# L'obsolescenza tecnologica

- Incompatibilità di versioni differenti di HW e SW
- Progressiva perdita delle conoscenze necessarie a gestire le vecchie versioni di HW e SW
- Evoluzione dei formati (es. .doc, .pdf, .mp3, etc.)
- Decadimento fisico dei supporti di archiviazione e memorizzazione

Indifferenza o scarsa sensibilità al problema da parte delle aziende, del legislatore, degli amministratori e degli *stakeholders* in genere

# Quanta informazione?

- HOW MUCH INFORMATION 2003?

<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/in>

Dati riferiti al 2002

- 800 Mb prodotti per ciascun abitante della terra
- 30% di incremento annuo fra 1999 e 2002
- 92% immagazzinata su dischi rigidi in formato digitale
- 0,01% disponibile su carta

Eterogeneità → Standardizzazione, Interoperabilità,  
Cooperazione

# La conservazione digitale

## una definizione:

- L'insieme delle attività e degli strumenti che assicurano che i documenti informatici siano mantenuti accessibili, utilizzabili (leggibili e intelligibili) e autentici (univocamente identificabili e integri) nel medio e nel lungo periodo, in un ambiente tecnologico presumibilmente diverso da quello originario (M. Guercio, 2005)
- L'espressione " 'nel lungo periodo' non significa fornire una garanzia per cinque o cinquanta anni, ma piuttosto lo sviluppo responsabile di strategie che possano affrontare i cambiamenti costanti prodotti dal mercato dell'informazione"

(U. Schwens, H. Liegmann. Citati da S.Strathmann, *Digital Preservation: an introduction*, 2007)

# L'infrastruttura profonda della conservazione digitale

- un sistema nazionale e distribuito di archivi digitali collettivamente responsabili dell'accessibilità nel lungo periodo agli oggetti digitali, che sono istanze dell'eredità sociale, culturale ed economica di una nazione;
- il carattere di affidabilità degli archivi, che devono essere in grado di gestire con sicurezza le operazioni di stoccaggio degli oggetti digitali, le procedure tecniche di conservazione e quelle necessarie al mantenimento dell'accessibilità;
- la definizione di un processo di certificazione degli archivi, che sia prova della loro affidabilità;
- la definizione di un diritto-dovere degli archivi certificati di esercitare attivamente una funzione di salvataggio dell'informazione digitale a rischio di distruzione o abbandono

*Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.* Commission on Preservation and Access and the Research Libraries Group. 1996 <http://www.oclc.org/programs/ourwork/past/digpresstudy/final-report.pdf>

# L'infrastruttura profonda della conservazione digitale

## *Digital stewardship* (gestione responsabile)

“conservare l'eredità culturale digitale è qualcosa di più del processo tecnico di mantenimento dei segnali digitali nel tempo. E' un processo sociale e culturale, che implica la scelta di quali risorse conservare e in che forma; è un processo economico, che implica l'armonizzazione di mezzi limitati con obiettivi ambiziosi; è un processo giuridico, che implica la definizione del complesso dei diritti necessari alla conservazione permanente delle risorse culturali; è una questione di responsabilità ed incentivi, e di articolazione ed organizzazione di nuove procedure gestionali. E soprattutto è un'impegno continuativo, di lungo periodo, spesso condiviso e gestito in cooperazione da una molteplicità di istituzioni”

B.Lavoie, L. Dempsey. *Thirteen Ways of Looking at...Digital Preservation*. In: *D-Lib magazine*, luglio/agosto 2004 <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>

# Un progetto italiano: Magazzini Digitali

Assicurare la conservazione di lungo periodo delle risorse digitali ricevute dalle Biblioteche nazionali centrali per deposito legale, e, più in particolare, di garantire nel tempo la loro:

- **vitalità (viability):** "capacità di sopravvivenza" delle sequenze di bit che compongono i file, che devono rimanere intatte e accessibili da parte di un elaboratore
- **traducibilità (renderability)** da parte di un elaboratore: un determinato hardware e un determinato software sono in grado di gestire le risorse digitali depositate, come ad esempio visualizzare a video un documento in formato PDF
- **autenticità:** documentazione della certezza dell'identità e della integrità delle risorse depositate
- **fruibilità da parte delle comunità di riferimento:** ad esempio la disponibilità di servizi quali quelli proposti dal modello FRBR: trovare, identificare, selezionare, ottenere una risorsa digitale (o un insieme di risorse).

<http://www.rinascimento-digitale.it/magazzinidigitali.phtml>

# Un progetto italiano: Magazzini Digitali

- Infrastruttura HW/SW scalabile (rack di PC con dischi SATA)
- SW di base interamente Open Source
- 40Tb iniziali
- Ridondanza: 2 siti principali (BNCF, BNCR), 1 dark archive fuori linea (BNM)
- 3 data center ISO 27001 in aree geografiche differenti e gestiti da società differenti

# Gli Identificatori Persistenti (PI)

- gli Identificatori Persistenti : "etichette" che permettono di **identificare e localizzare** il documento con continuità
  - **assegnare al documento un identificatore univoco**, che sia potenzialmente stabile oltre il ciclo di vita del documento e del produttore del documento, e questo è un problema di natura tecnologica
  - **assegnare all'identificatore univoco il documento**, assicurando che quel documento sia sempre lo stesso, cioè che il suo contenuto non venga alterato, e questo è compito del produttore del contenuto

(Norman Paskin, 2004)

# Gli Identificatori Persistenti (PI)

Modello "organizzativo" generale di un sistema di identificatori persistenti

- Selezione dei documenti digitali a cui attribuire un PI
- Attribuzione al documento digitale di un PI e creazione di un registro dei PI
- "Risoluzione" del PI, cioè associazione al PI della URL del documento
- Mantenimento del registro che associa PI e URL (anche se la URL cambia si mantiene il legame con il PI)

# Gli Identificatori Persistenti (PI)

- **Il Progetto NBN** : un'infrastruttura di registrazione e risoluzione di PI e un registro nazionale di PI per **l'identificazione univoca e l'indirizzamento affidabile e durevole** delle risorse digitali
- NBN: Registro internazionale mantenuto da LC
- Sub-namespace italiano: **NBN:IT ISO 3166**

<http://www.rinascimento-digitale.it/nbn.phtml>

# Gli Identificatori Persistenti (PI)

un modello di applicazione di NBN (progetto FRD/MIBAC/CNR)

