



# LA CONSERVAZIONE A LUNGO TERMINE DEI DOCUMENTI ELETTRONICI: NORMATIVA ITALIANA E PROGETTI INTERNAZIONALI

Maria Guercio  
Università di Urbino  
m.guercio@mclink.it

6/04/01



# INTRODUZIONE

# PERCHÉ LA CONSERVAZIONE È UN PROBLEMA CRUCIALE

- E' il nodo cruciale dell'automazione in campo documentario perché richiede un cambiamento significativo rispetto alle attività tradizionali: la conservazione in ambiente digitale è una funzione attiva e continua nel tempo per la quale non c'è ancora sufficiente esperienza, né un'adeguata analisi concettuale.
- I tempi degli interventi per il mantenimento della memoria si sono accorciati.
- La diversificazione dei prodotti non consente soluzioni univoche.
- La fragilità dei supporti e la facilità nella manipolazione richiedono investimenti significativi in termini di controllo dei depositi e sicurezza

## ... E IRRISOLTO

- L'obsolescenza è un fenomeno irreversibile e ambivalente: per affrontarla non si sono ancora individuati metodi condivisi, univoci, regolamentati
- L'indifferenza del mercato (e qualche volta anche del legislatore e degli amministratori) per le esigenze conservative di lungo periodo è ampiamente testimoniata dall'assenza di standard internazionali, dall'insufficienza delle norme nazionali e delle procedure interne alle amministrazioni e, in Italia, dalle ridottissime risorse per la ricerca.

# LA CONSERVAZIONE DELLE MEMORIE DIGITALI. RISCHI E VINCOLI

- La fragilità dei supporti è il male minore: l'evoluzione continua determina l'obsolescenza di hardware e software e la necessità di intervenire sui documenti e sui sistemi documentari per:
  - trasmettere fedelmente i documenti nel lungo periodo
  - per disporre di nuove tecnologie di gestione e accesso

# IL BUCO NERO DELLA CONSERVAZIONE DEI DOCUMENTI INFORMATICI

- Per l'automazione avanzata dei sistemi documentari servono standard per la conservazione degli archivi informatici che garantiscano due obiettivi contrastanti:
  - autenticità (identificazione e integrità) e accessibilità generalizzata e di lungo periodo dei sistemi documentari (dei documenti e delle relazioni di contesto)
- Le ricerche e le sperimentazioni in corso non hanno ancora dato risultati definitivi. Le soluzioni tecnologiche e organizzative, supportate da una adeguata analisi concettuale, devono consentire:
  - memorizzazione ed accesso a basso costo
  - protezione contro l'obsolescenza
  - sicurezza del sistema e nel sistema
  - accesso per utenti non specialisti del settore
  - utilizzo integrato di supporti diversi

# LA CONSERVAZIONE RICHIEDE INTERVENTI PIANIFICATI NELLA FASE ATTIVA

- E' necessario (per contenere i costi e garantire i risultati) intervenire precocemente sin dalla formazione dei documenti e sulle modalità di conservazione e accesso e sulla documentazione dei programmi: gran parte delle informazioni che garantiscono l'accesso all'archivio e la verifica dell'autenticità sono disponibili solo nella fase attiva della gestione documentaria (es. dati sull'organigramma, il piano di classificazione, la relazione tra contesto documentario e contesto amministrativo, lo schema logico dei db, la documentazione delle applicazioni, ecc.).
- Sono a grave rischio tutti quei documenti che non si producono in un contesto organizzativo definito e, quindi, controllato: ad esempio la documentazione di natura scientifica e personale (corrispondenza che utilizzi sistemi di posta elettronica, materiali di studio messi a disposizione sui siti web) che è caratterizzata da:
  - uso di formati proprietari che hanno una vita breve e sono costosi da migrare
  - assenza di metadati standard di identificazione (ad esempio di profili standard tipici dei sistemi di registrazione delle pp.aa.)

# LA NECESSITA' DI LINEE GUIDA. LO STUDIO MOREQ (MODEL REQUIREMENTS FOR THE MANAGEMENT OF ELECTRONIC RECORDS)

- Finalità generale: fornire linee guida per l'utilizzo ottimale di ICT per la gestione informatica dei documenti
- Obiettivi specifici: indirizzare i progetti di automazione dei sistemi documentari in modo da garantire:
  - la corretta formazione dei documenti: identificare e gestire gli elementi costitutivi
  - il miglioramento della fruizione: sviluppare funzionalmente i requisiti di base del document management
  - concrete possibilità di conservazione





# LA NORMATIVA NAZIONALE

# LA DELIBERA AIPA 51/2000: REGOLE TECNICHE IN MATERIA DI FORMAZIONE E CONSERVAZIONE DEI DOCUMENTI INFORMATICI

- Le regole sono state deliberate ai sensi dell'art. 18 comma 3 del dpr 513/97 e prevedono l'intesa con l'amministrazione degli archivi di Stato. Comprendono:
  - Definizioni di documento informatico, formazione del documento, conservazione, formato, sicurezza, accesso e archivio
  - Requisiti
  - Formato
  - Gestione
  - Conservazione
  - Trasmissione
  - Accesso
  - Sicurezza

## LA FORMAZIONE DEL DOCUMENTO INFORMATICO NELLA DELIBERA 51/2000

- Il processo di generazione del documento informatico deve essere riferibile con certezza al soggetto pubblico o privato che lo abbia prodotto o ricevuto
- E' sottoscritto con firma digitale quando prescritto
- E' sottoposto sempre alla registrazione (di protocollo o altra forma di registrazione)
- Si applicano, naturalmente, le norme stabilite dal testo unico sulla documentazione amministrativa (dpr 445/2000)

## CONSERVAZIONE DEL DOCUMENTO INFORMATICO NELLA DELIBERA 51/2000

- Ordinata custodia di documenti informatici al fine di assicurarne l'integrità, l'affidabilità e la consultabilità nel tempo anche attraverso idonei strumenti di ricerca

## DEFINIZIONE DI SICUREZZA NELLA DELIBERA 51/2000

- L'insieme delle misure organizzative e tecniche finalizzate ad assicurare senza soluzione di continuità l'integrità, la disponibilità e la riservatezza dei documenti e degli archivi informatici

## DEFINIZIONE DI ARCHIVIO NELLA DELIBERA 51/2000

- L'insieme organizzato e gestito in modo unitario per aree omogenee di uno o più supporti (*sic*), univocamente identificati, contenenti i documenti registrati (*sic*)

# I FORMATI DEI DOCUMENTI INFORMATICI (ARTICOLO 4)

- Formato: le modalità di rappresentazione del contenuto
- Requisiti dei formati:
  - consentire l'archiviazione, la leggibilità, l'interoperabilità e l'interscambio
  - la non alterabilità del documento durante le fasi di accesso e conservazione
  - la possibilità di ricerca tramite indici di classificazione o di archiviazione (*sic*), nonché sui contenuti dei documenti
  - l'immutabilità nel tempo del contenuto e della sua struttura (è fatto divieto di includere nei documenti informatici macroistruzioni o codice eseguibile che introducano modifiche)
  - La possibilità di integrare il documento informatico con immagini, suoni e video purché incorporati in modo irreversibile

## LA CONSERVAZIONE DEI DOCUMENTI INFORMATICI (ART. 7)

- Si rinvia alla delibera Aipa 24/1998 (archiviazione ottica sostitutiva) e all'articolo 60 del dpcm 8 feb. 1999 (per l'estensione della validità anche ai documenti nati su supporto informatico)



## INSUFFICIENZA DELLE NORME PROPOSTE

- Le norme hanno un andamento non lineare. In alcuni casi si caratterizzano per la genericità delle definizioni e delle disposizioni, in altri il dettaglio tecnologico finisce per complicare e rendere inattuabile e incomprensibile l'applicazione (ad esempio nel caso delle regole sull'archiviazione ottica)
- Manca una struttura concettuale chiara, condivisa e coerente, della cui rilevanza non c'è consapevolezza
- Il legislatore ha avuto fretta di completare il quadro normativo e ha – erroneamente – ritenuto di poter affrontare la conservazione di lungo periodo come un "semplice" problema di migrazione e di ridurre il problema della verifica dell'autenticità nel tempo a un problema di autenticazione nella fase di formazione, trasmissione o riproduzione.

# LA CONSERVAZIONE DEI DOCUMENTI INFORMATICI. LO STATO DELL'ARTE

- Le ricerche e le sperimentazioni promettenti riguardano pochi progetti internazionali, che per fortuna sono reciprocamente coerenti: i progetti OAIS, InterPARES, NPACI
- In tutti questi progetti si assume la centralità del fondamento teorico e la necessità di una nuova e approfondita riflessione concettuale dei principi e delle funzioni documentarie e archivistiche, ma anche la contemporanea analisi critica degli interventi e dei metodi di conservazione finora proposti



# I METODI PER LA CONSERVAZIONE

# I METODI PER LA CONSERVAZIONE DEI DOCUMENTI INFORMATICI

- Conservazione delle tecnologie nel loro ambiente operativo
- Imitazione/Emulazione
- Migrazione

# CONSERVAZIONE DI TECNOLOGIA

- **Conservazione di tecnologia:** consiste nel conservare o riprodurre l'hardware e il software originari insieme al loro ambiente operativo. Prevede due opzioni:
  - **Mantenimento del vecchio ambiente tecnologico**
  - **Riproduzione del vecchio ambiente**
- E' un metodo ormai rifiutato dalle amministrazioni per l'enorme quantità (in crescita esponenziale) di tecnologia diversa da mantenere. Esistono poi limiti fisici invalicabili. Ha il vantaggio di "sembrare" facile e di semplificare le procedure di verifica dell'autenticità dei documenti.

# IMITAZIONE/EMULAZIONE DEL SOFTWARE

- Prevede più opzioni:
  - emulazione
  - creazione di macchine virtuali
  - ingegnerizzazione del software
- Consiste nella produzione di software specifico per la creazione di un nuovo ambiente informatico che consenta di utilizzare i documenti a fronte di una evoluzione o di una perdita dell'hardware o del software nativi. E' una sorta di "strato" tra il sistema operativo e i programmi applicativi che "simula" le proprietà originarie del precedente ambiente.

# IMITAZIONE DEL SOFTWARE. SVANTAGGI

- Non è un metodo molto usato per fini di conservazione a lungo termine: è molto costoso e impegnativo e l'evoluzione tecnologica rende sempre più difficile la scrittura di emulatori di software sempre più numerosi, sofisticati e complessi (centinaia di migliaia di combinazioni possibili di software applicativi), di cui non si hanno generalmente conoscenze sufficienti. Ripropone, inoltre, i difetti delle tecnologie originarie.
- Crea problemi notevoli per lo sviluppo di help desk per assistere l'accesso futuro: ad esempio nei sistemi in rete sono necessari sistemi di aiuto alla ricerca molto complessi e diversificati
- Implica la rinuncia alla possibilità di utilizzare tecniche e prodotti migliori.

# MIGRAZIONE

- Consiste nel trasferimento di oggetti da un ambiente hardware/software a un altro. Può implicare modifiche anche rilevanti nel flusso di bit che costituiscono il documento.
- E' costosa e presenta molti rischi di perdita di informazioni rilevanti all'adeguato accesso archivistico ai documenti.
- Prevede più opzioni per la gestione dei formati:
  - il mantenimento dei formati originari (ad esempio il mantenimento dei vecchi formati word della Microsoft)
  - la migrazione dei formati
  - la produzione di formati standard (*persistent object*)



## MANTENIMENTO DEI FORMATI ORIGINALI

- Il mantenimento dei formati originali consente la capacità diffusa di manipolazione e, quindi, accresce i rischi per l'integrità dei documenti
- Non sono garantite le funzioni di archiviazione permanente

## MIGRAZIONE DEI FORMATI

- Non risolve il problema dell'obsolescenza, perché è basata su soluzioni ad hoc e a termine
- Non è certa l'esistenza di formati di archiviazione
- Non elimina la difficoltà di gestire una molteplicità eccessiva di formati
- Non sono sotto controllo i cambiamenti dei formati medesimi

## MIGRAZIONE IN FORMATI STANDARD

- Consiste nella trasformazione di qualunque oggetto informatico in un modello rappresentato con linguaggio standard (ASCII o UNICODE), incapsulato in un insieme di metadati espressi con marcatori e reso possibile mediante l'uso di linguaggi di marcatura

# USO DI FORMATI STANDARD PER I METADATI E I DOCUMENTI: I LINGUAGGI DI MARCATURA

- I linguaggi di marcatura consistono in informazioni testuali (ASCII o Unicode) che forniscono indicazioni di natura strutturale e semantica sui materiali trattati (documenti, libri, ecc.). Il linguaggio più che promettente è XML (eXtensible Markup Language, che costituisce la ottimizzazione per il web dello standard ISO SGML-Standard Generalized Markup Language
- Sia i metadati (di contesto documentario e amministrativo) che i documenti vengono conservati in formati tecnologicamente indipendenti e per componenti funzionali

## MIGRAZIONE IN FORMATI STANDARD: VANTAGGI E LIMITI

- Poiché implica una sola migrazione, si riducono i rischi di autenticità e si minimizzano i cambiamenti
- Consente controlli specifici e procedure predefinite per la verifica dell'autenticità
- E' una tecnica coerente con la natura archivistica dei documenti (oggetti stabili e relazioni definite)
- Consente l'utilizzo di tecnologie più avanzate che migliorano l'accesso e la gestione del sistema
- Ha il limite (temporaneo) di non disporre ancora di applicazioni di mercato



# LA RICERCA INTERNAZIONALE

## LA RICERCA INTERPARES

- E' rilevante per la situazione in Italia (ma in generale per la comunità internazionale) l'obiettivo di creare una struttura di riferimento concettuale e metodologica generale (definizione degli elementi che garantiscono l'identificazione del documento, definizione delle procedure per la verifica dell'integrità, per la selezione e per la conservazione da parte delle istituzioni archivistiche) coerente con la tradizione nazionale e qualificata

# LO STANDARD OAIS PER LA CONSERVAZIONE DEI DOCUMENTI INFORMATICI

- Il Reference Model for an Open Archival Information System (OAIS) è una proposta di standard ISO (in corso di approvazione) presentata per la salvaguardia dei sistemi di dati e gli archivi aerospaziali.
- Obiettivo: definizione di una guida metodologica alla conservazione permanente o a lungo termine di informazioni digitali
- Propone nuovi termini e nuove definizioni in quanto si tratta di una ipotesi di soluzione trasversale e comune a più settori disciplinari (archivi, biblioteche, centri di documentazione scientifici). Ha il limite di una genericità di approccio e una insufficiente attenzione alla continuità della funzione conservativa
- Si basa sul principio di considerare i documenti come entità cui associare informazioni di diverso tipo incapsulate in modo permanente nei documenti:
  - provenienza: storia del soggetto e della custodia dell'archivio
  - contesto documentario: dati relativi al vincolo archivistico
  - stabilità dei supporti
  - modalità ricerca



# IL PROGETTO XML. I PRESUPPOSTI TEORICI E TECNOLOGICI DELLA RICERCA

- All'origine c'è la convinzione che i documenti elettronici possano essere considerati come fonti distribuite di informazione *semi-strutturata*, costituite da uno schema definito di componenti informative interne ed esterne al documento e da una serie di elementi passibili di variazione (il supporto, il contesto tecnologico, ecc.).
- Il progetto americano si basa su una serie di presupposti e pre-condizioni:
  - la codifica ASCII o Unicode per le informazioni testuali e la codifica bitmap per le immagini sono indipendenti da infrastrutture tecnologiche,
  - la rappresentazione di informazione strutturata mediante linguaggi di marcatura (XML) è indipendente, di facile accesso e consente l'auto-descrizione di docc.,
  - la definizione di una metodologia per la creazione di fonti informative sostitutive degli originali è basata sullo sviluppo di "contenitori" (*wrapper*) di prodotti software strutturati in modo che:
    - tutti i metadati che descrivono i contesti documentari abbiano la forma di documenti XML forniti di specifiche DTD,
    - tutte le informazioni testuali siano convertite in documenti XML e le immagini in bitmap,
  - tutti i riferimenti a immagini e ad altri documenti all'interno di un documento archivistico siano convertiti in collegamenti permanenti a loro volta rappresentati in un formato XML compatibile.

# XML PER LA CONSERVAZIONE DEGLI ARCHIVI INFORMATICI: IL PROGETTO NARA-NPACI

- Il progetto si occupa di tre grandi classi di documenti elettronici (documenti testuali, documenti composti, documenti GIS) il cui accesso richiede l'uso di strumenti software. Il nodo centrale della ricerca, che corrisponde alla questione di fondo della conservazione delle memorie digitali, è quello di:
  - definire un meccanismo per la creazione parzialmente automatica della rappresentazione digitale dei documenti in forme indipendenti dal software e sostitutive di originali,
  - predisporre un prototipo di strumento software indipendente dalle piattaforme, sufficientemente robusto, flessibile e scalabile basato sull'utilizzo di XML in quanto standard emergente (e promettente) per la rappresentazione e lo scambio informatico sul web. La scalabilità dei prodotti riguarda la capacità di rispondere anche alle esigenze di depositi archivistici di medie e piccole dimensioni.

# L'ARCHIVI NAZIONALE DI WASHINGTON: UN GATEWAY DEI PROGETTI DI RICERCA INTERNAZIONALI

- Programma ERA ([www.nara.org/era](http://www.nara.org/era)): ospitalità agli interlocutori più vivaci nel campo della gestione e della conservazione dei documenti informatici
- Sostegno alle iniziative di cooperazione internazionale

## E IN ITALIA?

- Partecipazione della comunità archivistica (Università di Urbino, Direzione generale archivi e ANAI) ai progetti InterPARES e NPACI-SDSC
- Studio Aipa sull'interoperabilità: l'interoperabilità presenta criticità e difficoltà simili a quelle della conservazione, almeno per quanto riguarda la trasmissione dei documenti e la loro tenuta (sistemi proprietari e indipendenza dalle piattaforme tecnologiche)

Le amministrazioni competenti concentrano ancora l'attenzione e le risorse quasi esclusivamente sulla formazione dei documenti informatici e sulla costruzione di sistemi per la gestione integrata di flussi documentari e amministrativi e dimostrano un sostanziale disinteresse per i rischi di perdita della memoria documentaria informatica.

## QUALCHE INDIRIZZO

- <http://www.interpares.org> (progetto InterPARES)
- <http://www.ispo.cec.be/ida> (progetto Moreq)
- <http://jitc.fhu.disa.mil/recmgt/> (Standard 5015.2 "Design Criteria Standard For Electronic Records Management Software Applications" US Department of Defense)
- <http://ccsds.org/RP9905/RP9905.html> (proposta standard OAIS)
- <http://sdsc.edu/NARA> e <http://www.npaci.edu/DICE/Pubs> (progetto del National Archives di Washington e del Supercomputer Center di S.Diego)