



L'archivista d'impresa
Corso di Formazione - Modulo Avanzato
Le sfide del digitale

05

Long Term Storage
Garantire la conservazione
storica del documento digitale

Andrea De Calisti

Martedì 10 settembre 2013
ore 9.30 - 18.00

Assolombarda
Via Pantano, 9 - Milano

Indice dell'intervento

- Analisi del problema
- Tipi di tecnologie possibili
- Il processo conservativo
- Domande e “risposte”

Analisi del/i problema/i

Il problema principale: «semplificare»

Cosa debbo conservare ?

Lo diamo per scontato per ora-----?

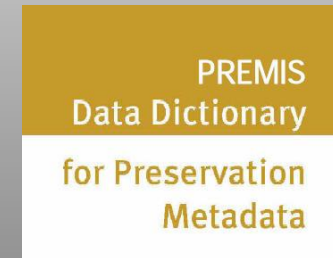
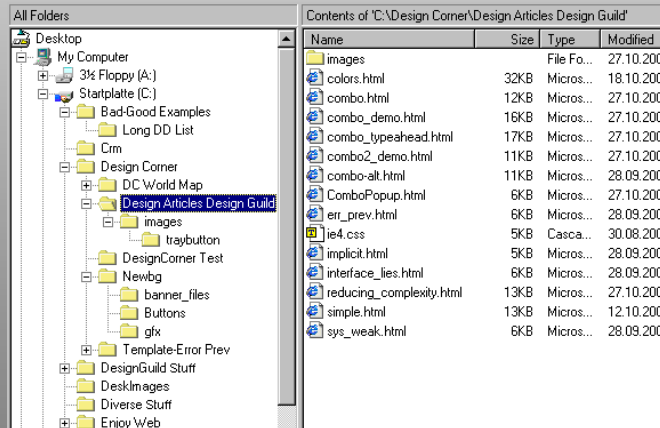


Una sequenza di bit più o meno ordinati

OVVERO

???????

La possiamo girare più o meno come vogliamo... ma alla fine....



La descrizione di ciò che voglio conservare... forse... e come intendo conservarlo

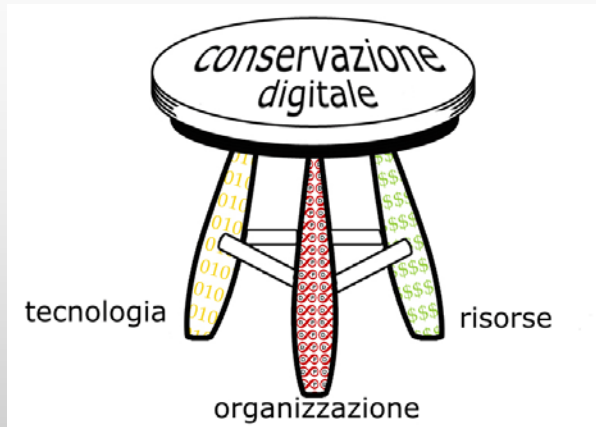
Scopo della conservazione ?

1) Avere la possibilità di recuperare un oggetto digitale con la certezza che questo oggetto sia assolutamente inalterato rispetto al momento in cui è stato immesso nel nostro processo di conservazione

2) Avere la possibilità di mantenere le informazioni SUFFICIENTI, a livello:

- Descrittivo dell'oggetto
- Utili a trovarlo
- Inerenti lo stato della conservazione, tipo:
 - Controllato
 - Migrato su nuovo supporto ecc ecc.

Di cosa abbiamo bisogno ?



Dice il saggio cinese... «sgabello regge se ha tre gambe».....

E ALLORA E' FACILE...VERIFICHIAMO SUBITO SE ABBIAMO TUTTE QUESTE COSE NO ??? E SE QUESTE COSE «POSSONO» CREARE DEI PROBLEMI A CUI DOBBIAMO PENSARE



Tecnologia.....è ormai decisamente a buon mercato...ma quali problemi pone ???

- Quali apparecchiature scelgo ???
- Quali supporti di conservazione scelgo ?
- Quali formati ?
- Quando cambio le apparecchiature ?
- E i supporti ?
- E i formati ?

Quante centinaia di volte avete sentito queste domande ? Domande importanti ma con risposte tutto sommato semplici...perché ??? E dopo lo vediamo...

Intanto ve le faccio io delle domande...se possibile....

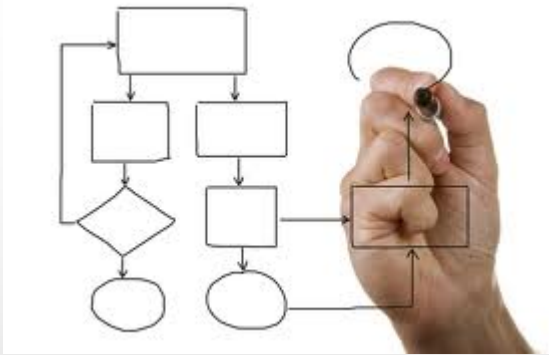


Risorse.....beh ok...commisuriamole

- Quante persone servono per mandare avanti un sistema di conservazione ?
- Servono delle professionalità particolari ?
- E' necessario aggiornarsi molto spesso ?

Quante centinaia di volte avete sentito queste domande ? Domande importanti ma con risposte tutto sommato semplici...perché ??? E dopo lo vediamo...

Intanto ve le faccio io delle domande...se possibile....



Organizzazione.....Questa si che ci piace....

Un adeguato processo organizzato può far risparmiare sulla tecnologia e sulle risorse....

E allora visto che siamo qui anche per questo proviamo a dare una possibile soluzione che sia sufficientemente rispettosa dei corretti criteri tecnici e metodologici

CONDIG...Una possibile soluzione

Prima cosa...il supporto prescelto

Vantaggi/svantaggi del nastro magnetico LTO

Vantaggi:

- Ottima sicurezza in termini di conservazione “statica”
- Ottimo rapporto prezzo/GB
- Ottimi sistemi di controllo della qualità dei dati e del supporto
- Ottimo rapporto spazio necessario allo stoccaggio/GB
- Possibilità di utilizzare librerie per la gestione automatica dei supporti

Svantaggi:

- “estrema” lentezza del trasferimento dati
- Programmi di scrittura/rilettura non compatibili
- Breve storia del supporto



Fattore di sicurezza dei diversi supporti

La valutazione è stata fatta su una scala da 1 a 10 dove 10 è il punteggio migliore. Le fonti sono, documentazione varia ed esperienza diretta

	Nastri magnetici	Supporti ottici	Supporti magneto ottici	Hard disk esterni/interni
Durata del supporto fisico	8	4	6	4
Protezione dei dati	9	3	8	4
Strumenti di recupero	5	5	5	8
facilità di conservazione	8	4	7	3

Costi per la produzione dei supporti

La valutazione è stata fatta su una scala da 1 a 10 dove 10 è il punteggio migliore. Le fonti sono, documentazione varia ed esperienza diretta

	Nastri magnetici	Supporti ottici	Supporti magneto ottici	Hard disk esterni/interni
Costo del supporto	0,02/GB	0,125/GB	0,18/GB	0,075/GB
Costo della produzione	8	2	6	9
Costo del controllo	5	2	7	9
Costo della conservazione del supporto	8	4	6	7

Due piccole testimonianze

http://www.youtube.com/watch?v=SvMehW_-5g0

[Perché un nastro LTO risulta essere migliore di un HD su uno scaffale](#)

<http://www.youtube.com/watch?v=X82sC97yQeE>

[LTFS su LTO, una soluzione per il recupero dati e il workflow](#)

LTFS questa «sconosciuta»

Cosa significa e cos'è LTFS:

Linear Tape File System (LTFS) refers to both the format of data recorded on [magnetic tape](#) media and the implementation of specific software that uses this data format to provide a [file system](#) interface to data stored on magnetic tape. The Linear Tape File System format is a self-describing tape format developed by [IBM](#) to address tape archive requirements. The *LTFS Formats* specification, which was adopted by the [LTO Technology Provider Companies](#), defines the organization of data and [metadata](#) on tape - files stored in hierarchical [directory structure](#). Data tapes written in the LTFS Format can be used independently of any external database or storage system allowing direct access to file content data and file metadata. This format makes it possible to implement software that presents a standard file system view of the data stored in the tape media. This file system view makes accessing files stored on the LTFS formatted media similar to accessing files stored on other forms of storage media such as disk or removable flash drives.

<http://www.youtube.com/watch?v=XT4eDgQ-fMg>
[LTFS – installazione e configurazione](#)

Abbiamo la tecnologia,
scopriamo il processo...o
meglio...un possibile
processo !

Processo di conservazione

- Fase di analisi/caricamento
- Fase di generazione dell'impronta digitale
- Fase di migrazione sui supporti LTO
- Fase di controllo
- Creazione della ridondanza
- Migrazione su nuovi supporti
- Consegnna degli oggetti digitali

Fase di Analisi/caricamento

I problemi da affrontare:

- **Analisi relativa alla salute dei dati**

I dati possono essere presi in carico ma chi assicura al fornitore del servizio di conservazione che i dati sono in buona salute ?

- **Sono dati strutturati ?**

I dati hanno un qualche tipo di ordine ? Alfabetico, archivistico ecc.

- **I dati sono rintracciabili ?**

I dati sono indicizzati ?

Fase di generazione dell'impronta digitale

Ma cos'è un impronta digitale ?

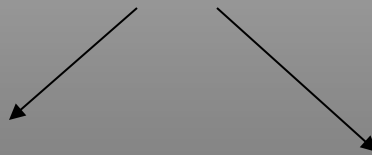
Il CRC/MD5 effettua un calcolo matematico su un blocco di dati (cioè un file), restituendo un numero caratteristico che rappresenta il contenuto e l'organizzazione stessa dei dati sottoposti a elaborazione. Lo scopo consiste nell'ottenere un valore numerico che identifichi univocamente quel particolare blocco dei dati. Si può pensare al CRC/MD5 come a una operazione capace di determinare l'impronta digitale di un file. Tale valore numerico, o impronta digitale che dir si voglia, è usato per identificare univocamente il file stesso ed è denominato "checksum" (valore di controllo o semplicemente "controllo").

Fase di migrazione sui supporti LTO

Gli oggetti digitali vengono memorizzati su nastri LTO in doppia copia, ovvero:

Ogni oggetto digitale risiederà su due diversi supporti. Durante questa fase saranno attivati tutti gli strumenti di controllo (sia hardware che software) di cui sono dotate le librerie di caricamento dei nastri LTO.

I nastri verranno conservati in due luoghi diversi



Fase di controllo

Ogni sei mesi i supporti vengono controllati e le impronte digitali confrontate con la lista originale



In caso di difformità su file l'operatore viene avvisato. Si effettua un controllo ed una eventuale copia del file "corrotto" dal file coerente con l'originale.

I supporti vengono controllati, tramite gli strumenti software a disposizione, anche dal punto di vista integrità fisica del supporto

I nastri vengono conservati secondo le specifiche tecniche messe a disposizione dalla casa

Creazione della ridondanza

Primo anno 2 nastri



Secondo anno 4 nastri



Terzo anno 6 nastri



Quarto anno:
eliminazione dei primi
due nastri con
sostituzione di due nuovi



Migrazione sui nuovi supporti

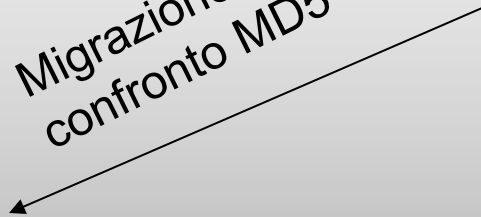
LTO tipo 4



Migrazione con
confronto MD5



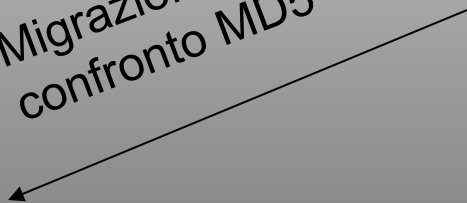
Migrazione con
confronto MD5



LTO tipo 5



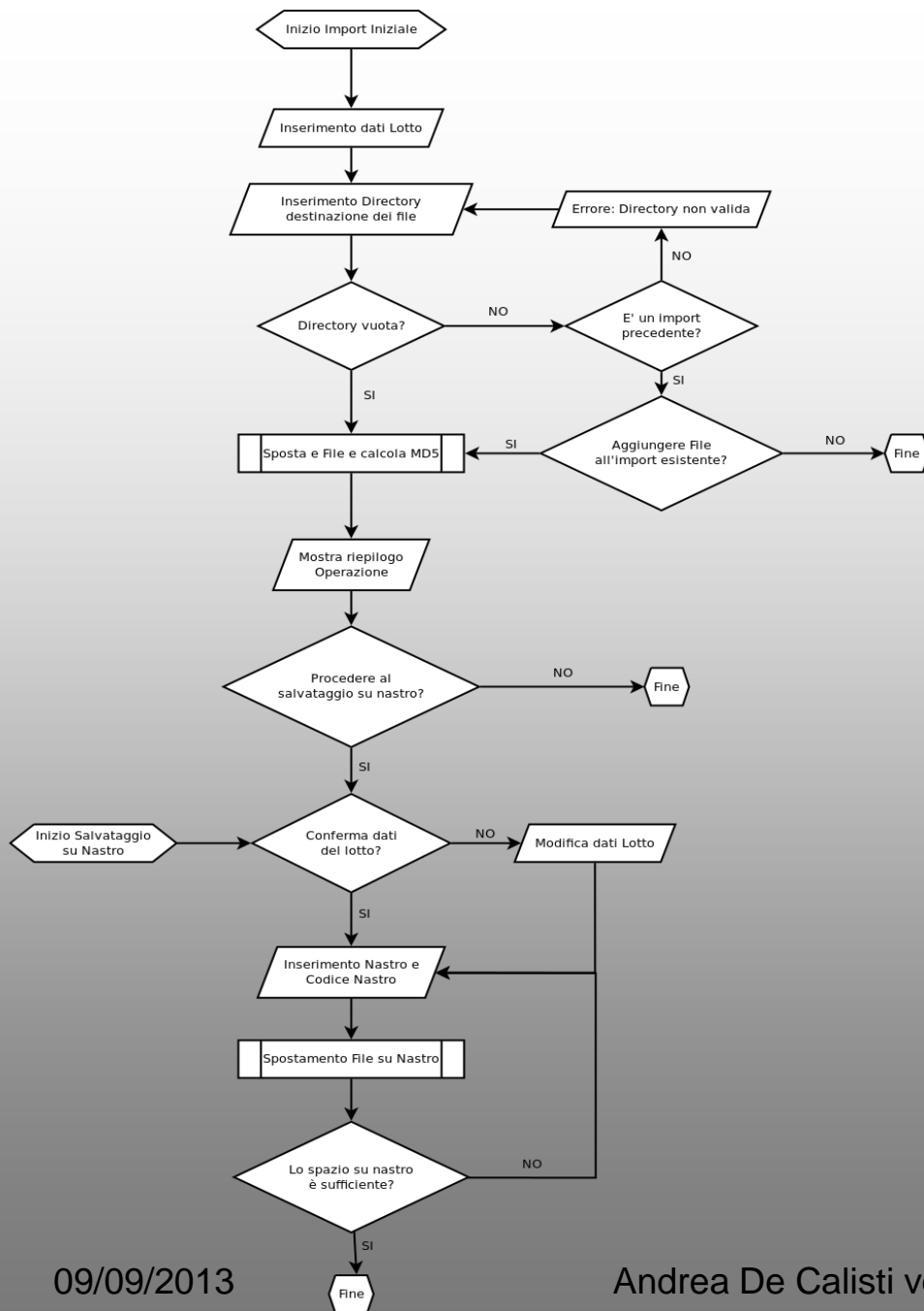
Migrazione con
confronto MD5



LTO tipo 6



- La nuova tecnologia viene assicurata dal fornitore di servizi o dal responsabile della conservazione
- I controlli MD5 vengono forniti al cliente



Il processo di importazione viene strutturato in lotti ed eventualmente in sublotti. Per ogni lotto si inseriscono le informazioni che si ritengono più utili (PER VOI !)...

- Data di creazione
- Responsabile del controllo
- Nome del lotto
- Derivazione lotto ecc ecc.

Un file sistema debitamente strutturato, nella maggior parte dei casi è già «parlante». Spendiamo le nostre energie su quello....

Un workflow ben studiato...è una possibile soluzione poiché contiene tutti i metadati necessari al recupero dei documenti....!

Cosa si ottiene

- Supporti nuovi ogni anno
- Controlli periodici
- Spedizione del log dei controlli
- Sostituzione del file difettoso e rigenerazione del supporto integrale
- Predisposizione e migrazione del servizio sulla nuova piattaforma tecnologica(LTO 6 ecc.)

“Conclusione”

Non esiste il supporto “perfetto”

Ogni caso va esaminato singolarmente

Esistono comunque delle “linee guida” e dei fatti oggettivi su cui basarsi

La conclusione finale per una «corretta» conservazione non può essere basata (per ora) su un supporto ma dalla combinazione del giusto supporto e del giusto processo “metodologico”

Domande ?