

**FONDAZIONE DELL'ORDINE INGEGNERI  
DELLA PROVINCIA MILANO**

**23 NOVEMBRE 2005  
SEMINARIO**

**GLI ARCHIVI INFORMATICI:  
PROBLEMATICHE DI INTEGRITA' NEL TEMPO.**

**Commissione per l'Ingegneria dell'Informazione  
Ing. Enrico P. Mariani**

**L'organizzazione degli archivi digitali e il costo relativo.**



## 1) Introduzione

I Francesi chiamano i computer "ordinateur": ordinatori, i Tedeschi li chiamavano cretini ad altissima velocità. Sta a chi li utilizza fare sì che al proprio computer si adatti la prima o la seconda definizione.

La lentezza e l'alto costo delle memorie di massa dei tempi passati (nel 1970 un disco da 128 kbytes costava quanto una Rolls Royce) costringeva gli addetti ai lavori ad ottimizzare lo spazio occupato e quindi a salvare solo i dati importanti, a compattarli, a codificarli in modo da evitare ridondanze, in una parola ad organizzarli.

Di contro l'ampia disponibilità di archiviazione ha portato gli utenti ad archiviare di tutto, spesso in modo ridondante, "tanto basta comprare un altro disco, per quello che costa".

In effetti ora i dischi più comunemente in uso hanno capacità di decine di Gigabytes, ed l'ordine di grandezza del costo è sceso da 1 Euro per 10 byte a 1 Euro per Gigabyte, quindi il costo degli archivi digitali pare diventato irrilevante. La realtà è diversa.

## 2) Il costo degli archivi digitali

Il tempo per accedere ai dati, e di conseguenza il costo della loro gestione, in assenza di una efficace organizzazione degli stessi, aumenta in modo impressionante con l'aumento del loro volume.

Se si hanno archivi ridondanti, per esempio più copie o versioni successive dello stesso documento non ben catalogate e sincronizzate, si rischia di utilizzare, modificare od inviare un documento obsoleto o con informazioni errate. In alcuni casi questo può causare gravi danni.

Più gli archivi sono voluminosi e sparsi, più difficile è pianificare una corretta strategia di backup. Infatti i dati organizzati richiedono normalmente meno spazio, di conseguenza l'hardware di backup sarà meno costoso, i backup saranno più veloci, sarà più facile riprenderli dalle copie.

Quanto detto sopra sembra banale, ma il caos nella maggior parte dei PC ormai è tale che uno dei prodotti sui quali investono ora le principali case di software è un motore di ricerca per il proprio PC.

Riassumendo:

L'organizzazione degli archivi è un costo in termini di tempo, sia in fase di progettazione delle modalità di archiviazione che nell'operatività quotidiana.

Il costo viene ripagato abbondantemente in fase di reperimento e utilizzo dell'informazione.

Più numerosi sono gli archivi più grande è il ritorno dell'investimento.

Per ridurre i costi di gestione degli archivi digitali occorre organizzarli.

### 3) L'organizzazione dei dati

Organizzare i dati digitali significa:

1) allocare i dati e le copie di backup sia dal punto di vista logico che fisico in modo che siano:

- Facilmente reperibili
- Facilmente utilizzabili
- Facilmente aggiornabili
- Disponibili solo agli utenti autorizzati
- Disponibili e leggibili nel tempo (almeno finchè sono necessari).

2) Proteggerli

1. Da intrusioni o danneggiamento intenzionale
  - furto di informazioni
  - alterazione o distruzione dei dati
2. Da distruzione accidentale
  - per danneggiamento o distruzione del supporto o dell'hardware
  - per obsolescenza dell'hardware
  - per obsolescenza del software o del formato
  - per errori del programma o dell'utente

Quando si pianifica l'organizzazione logica e fisica degli archivi (files) bisogna tener conto che il dati digitali non sono tutti uguali ma differiscono tra di loro per:

- dimensioni
- software necessario al loro utilizzo
- necessità di conservazione nel tempo
- modalità ed esigenze di utilizzo da parte degli utenti
- problemi di riservatezza
- criticità in caso di perdita, etc.

In base a queste caratteristiche si dovrà decidere:

- Dove allocare gli archivi (e.g. nel mio PC in area temporanea, nel mio PC in area che viene salvata, in un server, in un DVD, su una chiavetta USB ecc.)
- Come organizzare le cartelle che li contengono
- Come organizzare i backup
- Come organizzare gli accessi
- Se è necessario strutturare i dati in un data base.

Ecco alcuni suggerimenti pratici:

- Nel preparare un documento tenere le eventuali versioni intermedie in un area temporanea e distruggerle quando si è sicuri del documento definitivo
- Stampare i documenti importanti per rileggerli prima di archiviare la versione definitiva
- Usare per la versione definitiva i formati che garantiscono più stabilità nel tempo e rispetto al software (.txt, .pdf, .jpg, .odt).
- Evitare di usare caratteristiche apparentemente interessanti degli applicativi che potrebbero essere sensibili ai cambiamenti del software. (e.g. Da un certo momento in poi MS EXCEL non ha più fornito le versioni con i comandi in lingua, e tutti i fogli con i comandi in Italiano o Francese sono diventati obsoleti)
- avere un unico documento master. Se un file deve essere archiviato per scopi diversi evitare di duplicarlo, ma creare un sistema di indici.
- Non archiviare mai files con oggetti linked. Un documento dinamico non è un documento affidabile nel tempo. Gli oggetti linked vanno benissimo come strumento di lavoro per arrivare al documento definitivo, ma questo deve contenere i dati (embedded).

L'organizzazione dei dati comprende anche la gestione della loro protezione da intrusioni o danneggiamento intenzionale o accidentale.

i pericoli principali sono i seguenti:

- Furto dell'hardware  
E' importante proteggere il proprio hardware dai furti perchè quando rubano un computer, o magari come solo il disco fisso, il danno più grande per il derubato è la perdita dell'informazione, la quale talvolta può essere il vero obiettivo del ladro.  
I furti più insidiosi, perchè magari non ci se ne accorge, possono essere quelle di una copia ormai inutilizzata (vecchi backup, copie temporanee per trasferimento dati), di nessun valore per il derubato, ma piene di informazioni preziose per il ladro (disegni, documenti, programmi).
- Dati sottratti da hardware prestato, venduto o mandato in riparazione.  
Prima che un computer lasci il nostro studio per qualunque motivo, assicurarsi che sia pulito da ogni nostro contenuto di informazione. In caso di vendita formattare il disco fisso.
- Accessi non autorizzati. Per evitarli adottare le seguenti precauzione:
  - Proteggere fisicamente l'hardware da accessi indesiderati (chiudere a chiave).
  - Avere una politica di password proporzionata ai contenuti da difendere ed assicurarsi che venga applicata.
  - Utilizzare, ove possibile, formati pubblici.  
Evitare di creare documenti modificandone uno precedentemente creato: potrebbe contenere informazioni invisibili al momento, che non si vogliono mandare in giro. Ci sono esempi passati alla cronaca di documenti resi disponibili al pubblico con omissis abbastanza facilmente leggibili con un minimo di conoscenze di software.
- Accesso non sicuro ad Internet  
La propagazione di virus e spyware avviene sfruttando vulnerabilità dei browser, dei sistemi operativi e dei sistemi di posta. Esistono software più o meno soggetti ad attacchi, sia per la loro diffusione che per debolezze intrinseche.  
E' utile scegliere, quando possibile, i software più difficilmente attaccabili ed è necessario applicare le patch (pezze) che risolvono i problemi di sicurezza con la massima tempestività.  
Inoltre è bene configurare il browser nel modo meno attaccabile ed è indispensabile utilizzare firewall, software antispyware e antivirus e ricordarsi di aggiornarli.
- Utilizzo di congegni wireless (Wi-Fi, bluetooth)  
Se utilizzati con leggerezza sono una delle fonti di facili furti di informazioni, per esempio di rubriche telefoniche di VIP. Per proteggersi è importante:  
Cambiare la password (moltissimi lasciano quella default) e, se del caso modificarla periodicamente.  
Se possibile autenticare le macchine sulla rete.

Per quel che riguarda la distruzione accidentale dei dati è fondamentale avere una corretta politica di copie di sicurezza (backup).

Le cause principali di perdita accidentale dei dati sono:

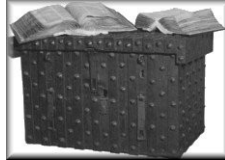
- Errore dell' operatore  
E' il caso più frequente, basta un comando errato. Non è grave se ci si accorge subito e si hanno le copie in ordine e rapidamente ripristinabili.  
Una norma fondamentale è quella di fare una copia dei dati interessati, magari semplicemente sul proprio PC, prima di eseguire operazioni critiche, quali aggiornamenti o conversioni di dati.

- Obsolescenza dell'hardware o del software

E' un problema grave, del quale si ha poca percezione.

Tutti sappiamo che è ormai quasi impossibile, leggere le cartucce degli anni 70 e 80, i floppy da 8 pollici e da 5 ½, difficile leggere quelli da 3 ½ che ci hanno tenuto compagnia per anni. Nessuno ci garantisce che tra qualche anno saranno ancora leggibili tutti i tipi di CD o DVD.

Uno degli esempi più famosi di perdita di informazione per obsolescenza è il progetto Domesday della BBC.



Il Domesday book fu scritto nel 1086 per incarico di Guglielmo il Conquistatore, e descrive in due volumi, ancora oggi consultabili, il catasto e le risorse dell'Inghilterra di allora. La BBC ha investito 2.5 milioni di sterline per creare una versione multimediale del Domesday Book che fotografasse lo stato della nazione nel 1986. A 16 anni di distanza tutto questo materiale, equivalente a parecchie enciclopedie, è illeggibile.

Chi volesse saperne di più può fare riferimento all'articolo di "The Observer"

**Digital Domesday Book lasts 15 years not 1000 di**

**Robin McKie and Vanessa Thorpe**

**Sunday March 3, 2002**

E' quindi necessario, quando si cambiano macchine, sistemi operativi o applicativi, verificare la compatibilità hardware e software dei nostri archivi e procedere, se necessario, alla conversione. Bisogna inoltre tenere presente che ogni conversione può portare perdita di dati o di caratteristiche del documento. Più i documenti utilizzano caratteristiche non standard, più il rischio è alto.

- Malfunzionamento, rottura o distruzione fisica di parte o di tutto il sistema informativo

Nei casi di perdita di dati il ripristino deve essere fatto dalle copie, per questo è necessario:

- Avere una politica di backup semplice, metodica e proporzionata alle necessità.  
Per esempio se una copia serve solo per il tempo di fare un'operazione critica è corretto farla sullo stesso disco di lavoro dell'originale. E' invece un errore imperdonabile usare lo stesso disco, magari su una partizione diversa per i backup di routine. Sono raccomandazioni che sembrano superflue, ma rispecchiano situazioni realmente viste sul campo
- tenere le copie di backup importanti fisicamente distanti
- Scegliere i formati dei documenti in modo da minimizzare il rischio di obsolescenza, possibilmente usare formati standard. Per i data base assicurarsi che siano leggibili dai data base manager più diffusi, questo requisito è ora facilmente ottenibile grazie allo standard ODBC.
- Per archivi veramente importanti (e.g. Ricevute INARCASSA) ricordarsi che l'archivio più sicuro e più facilmente leggibile è la carta.
- Controllare periodicamente le copie