

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

Corso di Laurea Triennale in Scienze di Internet

Dati digitali effimeri
e
permanenti

Relatore:
Chiar.mo Prof.
Davide Sangiorgi

Presentata da:
Massimiliano Canzonieri

Sessione I
Anno Accademico 2011-2012

*A mia mamma
a Katia
ai miei migliori amici
grazie per il vostro sostegno e
per essermi stati sempre vicini*

Indice

Introduzione	3
1 Il documento digitale	5
1.1 Introduzione	5
1.2 Cos'è un file	6
1.3 I formati	8
1.4 Il processo di digitalizzazione	12
1.5 Differenze tra analogico e digitale	14
1.6 Caratteristiche del documento digitale: vantaggi e problematiche	15
1.6.1 Flessibilità	16
1.6.2 Riproducibilità e conservazione	16
1.6.3 Trasmissibilità	17
2 Gestione dei documenti in ambiente digitale	21
2.1 La conservazione	21
2.2 Standard	22
2.3 Metadati	23
2.4 I codici in uso	25
2.5 Formati e metalinguaggio	26
2.5.1 SGML	28
2.5.2 XML	30
2.5.3 HTML	30
2.5.4 OAIS	31

3	Le memorie di massa	33
3.1	Classificazione delle memorie	33
3.2	Tipi di memorie di archiviazione	35
3.2.1	Memoria cartacea	35
3.2.2	Memoria magnetica	35
3.2.3	Memorie ottiche	39
3.2.4	Memorie elettroniche	42
3.2.5	Memorie olografiche	46
3.2.6	Memorie a singolo elettrone	47
3.2.7	Memorie organiche	48
3.2.8	HAMR	48
3.2.9	Patterned Media	49
3.3	Modi organizzativi della memoria	49
3.3.1	Accesso diretto o accesso sequenziale	51
3.3.2	Tracce e settori	52
3.3.3	Velocità	53
3.4	Porte di comunicazione	54
3.4.1	Seriale o Parallelo	54
3.4.2	USB e Firewire	55
3.4.3	Wireless	56
3.5	Problematiche tecnologiche	56
3.5.1	Deterioramento fisico e software	57
3.5.2	Obsolescenza dei supporti	58
3.5.3	Obsolescenza dei formati	58
3.6	Strategie per la conservazione del formato digitale	58
3.6.1	Trasferimento sui supporti analogici	59
3.6.2	Conservazione tecnologica	59
3.6.3	Emulazione	60
3.6.4	Riversamento diretto (refreshing) e sostitutivo (migration)	60
3.6.5	Archeologia digitale	61

4 Backup e sicurezza informatica	63
4.1 Backup	65
4.2 Strategia di backup	66
4.3 NAS	67
4.4 SAN	67
4.5 Innovazioni, hard disk, server virtuali, Cloud storage	68
Conclusioni	72
Bibliografia	73

Elenco delle figure

1.1	Segnale digitale vs analogico	18
3.1	Floppy-Disk	37
3.2	Hard-Disk	39
3.3	Memorie removibili	45
3.4	Disco Rigido	46
3.5	Memorie Organiche	48
3.6	Cella di Memoria	50
3.7	Bit	51
3.8	Struttura Disco	53
4.1	Schema di un sistema cloud	70

Elenco delle tabelle

1.1	Tipologia file Web	8
1.2	Tipologia file Immagini	9
1.3	Tipologia file Audio	10
1.4	Tipologia file Video	11
1.5	Tipologia file Misti	12
3.1	Confronto tra i diversi supporti	42

Introduzione

Musica, video, fotografia, documenti, elaborazione dati, tutto ciò che ci circonda e che regolarmente usiamo è ormai diventato un flusso di numeri in sequenza, 0 e 1, acceso e spento, che vengono elaborati da processori sempre più potenti e veloci. L'era digitale ha ormai conquistato tutto ciò che poteva conquistare affermandosi in ogni dispositivo, portatile e non, che nel tempo ha assunto un potere molto più vasto di quello che ci si poteva immaginare. La realtà prodotta dalla tecnologia del digitale, e dell'informatica in generale, è una vera e propria rivoluzione che non riguarda soltanto un cambiamento nel modo di produrre le informazioni e distribuirle, ma che coinvolge soprattutto la società e l'iterazione degli individui, modificandone a volte anche la cultura. Ad oggi l'impressione che abbiamo è che sia tutto a portata di un click. Le informazioni sono sempre reperibili, in qualsiasi istante e in qualsiasi punto della terra semplicemente attraverso una connessione internet, e soprattutto destinate a rimanere per sempre accessibili nei vari supporti di memorizzazione. Tale percezione è talmente forte e radicata all'interno di ognuno di noi che difficilmente ci accorgiamo di quanto invece il sistema sia altamente fragile, e che nulla resta invariato. Le possibilità offerte dalla tecnologia sono molteplici e ad oggi possiamo usufruire di una quantità di supporti e di metodi ampi per proteggere i nostri ricordi, la nostra cultura e tutto ciò che è stato trasformato e creato nell'era digitale, ma di per se la sola memorizzazione sui supporti non è sufficiente a salvaguardare tutto questo nel futuro. La digitalizzazione è un processo tanto semplice quanto complesso che passa dalla codifica di una traccia analogica alla decodifica e rappresentazione nella sua forma digitale. Questo è uno dei primi se non l'aspetto più importante di tutto il processo in quanto, è proprio la codifica nel formato digitale che permetterà di avere una rappresentazione fedele dell'informazione che vogliamo digitalizzare. Maggiore sarà

il flusso numerico che caratterizzerà l'informazione digitale, maggiore sarà la sua fedeltà. Il flusso di dati viene poi manipolato, inglobato, trasformato dalle varie applicazioni software che utilizzeranno quelle informazioni. Più in dettaglio la tesi analizza i vari passaggi che portano alla digitalizzazione. I primi due capitoli infatti sono serviti a capire nel dettaglio lo scenario del mondo digitale. Più precisamente cosa si intende per digitalizzazione, cos'è una sequenza numerica, come viene suddivisa ed interpretata dai calcolatori e gli standard utilizzati per garantire la massima compatibilità con tutte le componenti hardware e software. La seconda parte, composta dai capitolo 3 e 4, analizza invece la tecnologia impiegata per la memorizzazione dei dati digitali. Le evoluzioni portate dalla tecnologia e come essa ha influenzato le varie tecniche di memorizzazione e di trasmissione dell'informazione digitale. Sono state analizzate in particolare tutte le memorie, cosiddette di massa, nelle quali memorizzare le informazioni, le loro caratteristiche, i loro usi, l'evoluzione che queste nel tempo hanno avuto e che ancora continuano ad avere, nonché le future memorie e le nuove tecniche di memorizzazione, che rivoluzioneranno sia gli attuali dispositivi che quelli futuri. Vedremo inoltre come il problema principale sia quello di conservare i dati nel tempo e soprattutto conservare la grande quantità di dati che ogni giorno viene prodotta e che pone l'attenzione sui limiti di tutti i supporti di memorizzazione, nonostante la continua evoluzione tecnologica e la sempre maggiore capacità di memorizzazione. Le tipologie di memoria utilizzate sono state suddivise e raggruppate in 6 categorie, ognuna delle quali contenente circa 3 tipologie diverse di memorie di massa contenute tutte nel capitolo 3: la prima categoria racchiude le memorie magnetiche la cui caratteristica risiede nella memorizzazione dei dati attraverso stati magnetizzati della materia; la seconda categoria comprende invece memorie composte da un sottile foglio metallico inciso da un laser; la terza categoria è composta dalle ultime evoluzioni tecnologiche caratterizzate da apparecchi elettronici denominati "memorie FLASH"; nella quarta categoria troviamo lo sviluppo di una tecnologia ormai prossima alla commercializzazione e che riguarda le memorie olografiche; nella quinta e sesta categoria possiamo trovare un accenno a tipologie di memorie totalmente rivoluzionarie che sono ancora in fase di sperimentazione ma che potrebbero letteralmente e radicalmente cambiare l'intero sistema di memorizzazioni dati e che si basano su memorie a singolo elettrone e memorie organiche. Infine sono state analizzate le tecniche

utilizzate per la conservazione dei dati attraverso l'utilizzo dei vari supporti e dei limiti e dei vantaggi nell'utilizzare l'una o l'altra tecnologia e il cambiamento che sta avvenendo anche in fase di storage attraverso l'utilizzo del cosiddetto CLOUD storage , una tecnica del tutto nuova di tutela e conservazione dei propri dati che, nonostante sia all'inizio, sta riscuotendo già un grosso consenso da parte di singoli cittadini ma anche delle grosse aziende. Scopriremo quindi che non sarà così scontata l'idea di poter usufruire sempre ed in ogni istante dei nostri dati e che, per quanto la tecnologia venga in nostro aiuto, la conservazione dei dati nel tempo dipenderà anche dal modo in cui noi li gestiremo.

Capitolo 1

Il documento digitale

1.1 Introduzione

La parola documento ha subito un profondo mutamento con l'avvento della tecnologia e del ventesimo secolo. Si è passati dal considerare il documento come qualcosa di tangibile e cartaceo, a qualcosa di immateriale, non fisico. Il valore stesso del documento e la sua utilità si sono evoluti passando dal pensare un libro come portatore “*cosciente*” di informazioni, informazioni trascritte per essere trasmesse nel tempo (ad *memoriam*), a materiale che indipendentemente da chi lo utilizza, impone volenti o nolenti, una data immagine di se stessi. Nei primi è possibile riconoscere elementi inconsci come il pensiero dell'autore, i tipi di colori utilizzati, la calligrafia e la personalità dell'autore, nei secondi invece è stata pienamente riconosciuta solo la moderna semiotica come portatori di informazione, da qui la definizione degli stessi in “neutri e silenziosi”. Per essere catalogati come digitali i documenti devono possedere una ben precisa caratteristica: la loro numerabilità. La digitalizzazione del documento infatti è un processo che deve essere considerato a parte rispetto alla memorizzazione sul supporto in cui il dato è registrato. La parola digitalizzazione (dall'inglese *digit*), pone le sue radici nel sistema contabile che è il più logico e naturale metodo utilizzato da tutte le culture in cui i sistemi decimali sono presenti. Le dita esprimevano simbolicamente numeri con i quali contare. Il termine *digit* esprime perciò un simbolo o un numero. Ecco che, quando si converte un documento tramite scanner, lo si converte in una sua rappresentazione numerica,

simbolica e immateriale fatta di numeri. Essendo una rappresentazione numerica la si può inserire all'interno di un'altra rappresentazione numerica più complessa con la quale è possibile interagire. Tutto è ricondotto ad un sistema binario di 0 e 1. Ogni singolo punto di ciò che vediamo su uno schermo del pc, ogni sua parte visiva non è altro che una sua rappresentazione e codifica in formato binario. Per poter però trasformare il mondo reale in mondo digitale bisogna che siano attribuiti il giusto numero di bit. Maggiore è il numero di bit che esso contiene al suo interno, maggiore sarà la possibilità di rappresentarlo nella maniera più realistica possibile. Se prendiamo ad esempio la rappresentazione dei colori, per ogni bit vengono utilizzati tre byte. Ogni byte assume fino a 256 differenti combinazioni di 0 e 1 che rappresentano tutte le possibili sfumature di un colore. Generalmente per la rappresentazione video si utilizzano sistemi RGB a 24 bit. Quindi moltiplicando 3 byte avremo 256^3 possibili combinazioni di sfumature e dunque 16,7 milioni di colori. Un sistema a 8 bit avrà solamente un byte e dunque unicamente 256 possibilità. Risulta chiaro come nei sistemi con una mole di bit molto più alta vi è la necessità di avere componenti hardware e software con enormi quantità di calcolo. Avendo la possibilità di tradurre in una sequenza numerica il mondo che ci circonda e potendo mettere il tutto dentro una sequenza di 0 e 1 più complessa, ci permette di utilizzare una caratteristica fondamentale dell'ambiente digitale, l'interattività. È un elemento innovativo che cambia radicalmente le regole che accompagnano il mondo digitale rispetto a quello analogico poiché permette di avere un controllo completo delle informazioni.

1.2 Cos'è un file

Il termine file (termine inglese "archivio") indica una collezione di dati digitali memorizzati su un disco che può essere letta solo da un software. Non vi è una unica organizzazione. Le possibilità che ci vengono offerte per memorizzare tali dati all'interno di supporti sono svariate. In generale il sistema operativo ci fornisce una rappresentazione dei dati tramite uno schema ben fissato (File System) di strutture fisiche (in genere ad albero), ma nella pratica ogni utente è libero di ripartire e organizzare i documenti in maniera libera con l'ausilio delle interfacce grafiche. Per cominciare i file sono rappre-

sentati e identificati da un nome e un cognome o per meglio dire dalla forma nominale nome.estensione. Per quanto riguarda la parte del nome essa è trascurabile in quanto è l'unica parte del file che anche se deve seguire determinate regole di composizione non rappresenta un fattore determinante per il file. La seconda parte invece, la cosiddetta estensione, ne indica invece la tipologia dati a cui appartiene quel determinato file. Spesso l'estensione è anche nascosta alla visualizzazione dell'utente in quanto la sua modifica può causare un difetto di lettura e di interpretazione da parte del software. Assume perciò un ruolo fondamentale per l'effettiva leggibilità e per il suo riconoscimento. Tipicamente i file sono organizzati in directory o cartelle all'interno del supporto di memorizzazione. I file memorizzati appartengono ad un particolare tipo di formato per i quali esistono diverse applicazioni che possono semplicemente far visualizzare il file oppure permetterne le modifiche. Dato l'alto numero di aziende che creano file, per garantire la massima usabilità, sono stati creati degli standard. Lo standard è un insieme di norme e specifiche, che rappresentano un punto di riferimento per la realizzazione di tecnologie tra di loro compatibili. Esistono diversi enti a livello internazionale che si occupano di fornire la documentazione esaustiva, la sintassi, la semantica, il contesto operativo e la modalità di utilizzo dei vari formati digitali.

- ISO (International Organization for Standardization);
- IEC (International Electrotechnical Commission);
- IEEE (Institute of Electrical and Electronics Engineers)
- ETSI (European Telecommunications Standards Institute);
- CEPT (Conferenza Europea delle amministrazioni delle Poste e delle Telecomunicazioni);
- ITU (International Telecommunication Union);
- IETF (Internet Engineering Task Force).

Oltre a garantire l'interoperabilità tra i sistemi un file digitale ha sicuramente tra i suoi scopi principali quello di preservare le informazioni, documenti e tutto ciò che è

soggetto all'usura del tempo e ai limiti della memoria umana. Se da un lato l'utilizzo dei file digitali semplifica notevolmente molte problematiche relative alla loro distribuzione, usabilità, conservazione, dall'altro lato l'uomo potrebbe avere grossi problemi nel percepire e comprendere i dati digitali qualora non fosse più possibile utilizzare le macchine di conversione.

1.3 I formati

Vi sono diverse categorie di file che presentano sostanziali differenze. A seconda che il file sia di tipo testo, audio, video, il documento viene sempre associato ad un prodotto software per la codifica e decodifica. Per i file appartenenti al web abbiamo:

Subtype	Estensioni
plain	txt, text
richtext	
enriched	
html	htm, html
sgml	sgml
xml	xml,dtd
rtf	rtf

Tabella 1.1: Tipologia file Web

In questa categoria appartengono tutti quei file che possiedono al loro interno dati codificati in forma testuale. Tra tutti i dati quelli che sicuramente hanno una codifica più semplice in formato digitale sono i testi. È l'unica categoria che può essere considerata come multicanale, può cioè essere sia letta su carta stampata che su uno schermo di pc, oppure può essere utilizzato un programma software di lettura. Per codificare e scrivere questa categoria di dati è possibile utilizzare un banale editor di testo come un blocco note, anche se è preferibile utilizzare sempre dei tool appositi per la corretta visualizzazione di tali codifiche. Questo tipo di file in genere sono interscambiabili tra applicazioni molto differenti tra di loro, sono facilmente trasmissibili sia via disco che via

rete in quanto essendo solo file di testo la loro struttura grafica è totalmente ignorata ed assente e si riduce al massimo la dimensione del file. Con l'avvento di internet in particolare, questi tipi di file hanno aggiunto man mano diverse proprietà che hanno permesso loro di evolversi in qualcosa che va oltre la semplice lettura di informazioni, i cosiddetti ipertesti¹, documenti digitali con caratteristiche multimediali quali immagini, video, audio. La maggior parte dei documenti digitali in circolazione appartengono a tipi di formati vecchi per le loro caratteristiche di codifica e decodifica. Ancora oggi sono largamente utilizzati grazie proprio alla loro compatibilità con molte piattaforme diverse che garantisce, di non dover ricorrere a software o dispositivi hardware di terze parti adatti per leggerli.

Subtype	Estensioni
tiff	tiff, tif
gif	gif
jpeg	jpeg, jpg, ipe, jfif, pipeg, pjp
x-xbitmap	xbm
png	png
x-ms-bmp	bmp
fractals	fif

Tabella 1.2: Tipologia file Immagini

In questa categoria appartengono immagini statiche. Possono essere visualizzate solo se in possesso di strumenti hardware e software come schede grafiche specifiche per la visualizzazione di immagini e schede video che supportano determinate risoluzioni. La conversione delle immagini da analogico a digitale avviene per mezzo di dispositivi come scanner o fotocamere digitali. Vi sono sostanziali differenze tra i vari formati in particolari tra quelli che gestiscono e manipolano i bit e quelli che gestiscono i formati

¹É una parola marcata (sottolineata o evidenziata), in maniera da renderla riconoscibile, e che rimanda ad un altro documento diverso presente sul web. <http://it.wikipedia.org/wiki/Ipertesto>

vettoriali². Per i primi basta attivare e disattivare singoli bit che compongono l'immagine sullo schermo per modificarne la gradazione di colore o la qualità dell'immagine. Solitamente vengono anche definiti come immagini rasterizzate³ e in questo caso si ha la cosiddetta perdita di informazione (l'immagine compressa è simile a quella originale ma non uguale). Le immagini di tipo vettoriali invece, vengono gestite come delle entità matematiche. In questo caso essendo una formula matematica il limite di visualizzazione sarà dato solo dal dispositivo di output. Sarà facile distinguere i due tipi di formati in quanto uno zoom mette in evidenza come su di una immagine formata da bit si avranno bordi seghettati, non lisci, in cui è possibile vedere con chiarezza il colore rappresentato dal singolo bit, mentre per le immagini di tipo vettoriale questa differenza non esisterà e si vedrà solamente una sequenza continua di luci e colori a qualsiasi livello di zoom riusciremo ad arrivare. Tra i più diffusi standard utilizzati troviamo sicuramente quelli jpeg e gif per immagini di tipo bit-mapped, mentre sicuramente il tif è lo standard più diffuso per le immagini vettoriali.

Subtype	Estensioni
basic	au, snd
x-aiff	aif, aiff, aifc
x-waw	waw
x-pn-realaudio	rm, ra, ram
mpeg	png
x-ms-bmp	mp2, mp3, mpa, abs, mpega
midi	Midi, mid

Tabella 1.3: Tipologia file Audio

A questa categoria appartengono tutti quei file sonori digitalizzati. Anche in questo caso il passaggio da un segnale analogico ad un segnale digitale viene attuato attraverso

²Sono immagini rappresentate da equazioni matematiche e sono indipendenti dalla risoluzione. Ingrandire una porzione della linea vettoriale non produrrebbe artefatti visivi o la visualizzazione dei singoli pixel http://it.wikipedia.org/wiki/Grafica_vettoriale

³Sono immagini formate da una matrice di pixel N-M in cui ogni pixel assume una profondità di colore specifica che definisce i vari punti dell'immagine <http://it.wikipedia.org/wiki/Rasterizzazione>

specifici strumenti come microfoni e schede audio e come per tutti gli altri formati digitali anche in questo caso si parla di memorizzazione dei dati senza perdita di informazioni o con perdita di informazioni. Comune a tutti i formati audio compressi vi è il formato PCM⁴ (Pulse Code Modulation) il quale campiona una onda sonora per un certo numero di volte al secondo in base ad una scala che può essere più o meno dettagliata, a seconda del tipo di segnale da misurare. Bisogna fare una sostanziale distinzione tra il formato e il codec utilizzato. Il codec è l'algoritmo che esegue la codifica e la decodifica dei dati in ingresso e in uscita, mentre il formato è il tipo di dato che evidenzia le caratteristiche del formato digitale. Questa differenza è sostanziale quando si utilizza un player multimediale che può o non può supportare determinati formati audio. Il formato che garantisce l'effetto finale migliore in riferimento a qualità e compressione è sicuramente l'mpeg layer 3 o come viene comunemente definito mp3. Discorso ben più diverso riguarda i formati midi per i quali il computer non deve eseguire dei suoni ma riceve una serie di comandi che comunicano al software che tipo di note suonare in termini di durata e ampiezza.

Subtype	Estensioni
mpeg	mpeg, mpg, mpe, mpv, vbs, mpegv
mpeg2	mpv2, mpv2
quicktime	qt, mov, moov
x-ms-video	avi
x-sgi-movie	movie

Tabella 1.4: Tipologia file Video

Anche per le immagini in movimento, come per l'audio devono essere utilizzati degli strumenti per convertire il segnale e renderlo digitale. Anche in questo caso ci sono numerosi formati di compressione ma anche qui si è riusciti a raggiungere un formato divenuto standard che unisce la qualità delle compressioni alla diminuzione di spazio utilizzato, il cosiddetto MPEG-2 che nonostante adotti una tecnica di compressione con perdita di dati mantiene alta la qualità finale dell'immagine. L'algoritmo di compressione

⁴È un metodo di rappresentazione digitale di un segnale analogico http://it.wikipedia.org/wiki/Pulse-code_modulation

lavora eliminando le informazioni ridondanti in ogni fotogramma, come scene uguali tra vari fotogrammi o informazioni non percepibili per l'occhio umano. L'elaborazione dei video avviene tramite software che possono lavorare sia con materiale analogico che digitale. Nel caso di questi tipi di file digitali, viene richiesta una quantità di spazio molto più grande rispetto alle altre categorie di file digitali.

Subtype	Estensioni
octet-stream	
post-script	ai, eps, ps
pdf	pdf
mword	doc, dot
vnd.ms-excel	xls, xlt, xlm, xld, xla, xlw, zll
vnd.lotus-1-2-3	123, wk4, wk3, wk1
vnd.musican	mus
vnd.ms-access	mdb, mda, mde
x-javascript	js, mocha
x-perl	pl

Tabella 1.5: Tipologia file Misti

Come possiamo vedere dalla tabella 1.5 in questo gruppo sono compresi tutti quei file che non trovano allocazione nelle precedenti categorie. Si tratta di formati binari il cui contenuto spazia nelle diverse categorie e che sono codificati da programmi proprietari per la cui visualizzazione e modifica è necessario acquistare il software. Ad esempio se vogliamo utilizzare un file .doc dobbiamo utilizzare il software proprietario di Microsoft, microsoft word. Oppure, come nel caso del documento PDF (Portable Document Format), bisogna utilizzare il software adobe per poter manipolare tale documento.

1.4 Il processo di digitalizzazione

Nel corso degli anni e con la diffusione di massa dei computer si è assistito ad una rapida evoluzione della sostituzione del documento cartaceo con quello digitale. Ar-

chiviazione, posta elettronica, fatturazione elettronica, rappresentano solo alcune delle forme di utilizzo del documento digitale che, nel caso dei documenti inerenti all'amministrazione pubblica, vengono ormai riconosciuti come autentici e sostitutivi del modello cartaceo. La digitalizzazione è soltanto una delle tante possibili rappresentazioni delle informazioni che il computer utilizza. Il sistema di numerazione scelto per rappresentare le informazioni sul computer è formato soltanto da due simboli, lo 0 e l'1. Per un sistema binario questi due simboli bastano per permettere di rappresentare qualsiasi informazione e di manipolarla a nostro piacimento. Il motivo che sta alla base della scelta di questa rappresentazione è fortemente legato al contesto in cui viene utilizzata questa rappresentazione numerica. Il computer infatti ha la capacità di comprendere solo un sistema formato da due stadi, 0 o 1, che corrispondono agli stadi elettrici acceso/spento. Con un solo bit si possono esprimere due stadi: 0, 1

Con due bit si possono esprimere quattro stadi: 00, 01, 10, 11

Con tre bit si possono esprimere otto stadi: 000, 001, 010, 011, 100, 101, 110, 111

La digitalizzazione è frutto di due processi ben precisi:

- Campionamento;
- Quantizzazione.

Il campionamento è un processo che, indipendentemente dal tipo di dato che vogliamo rappresentare sotto forma digitale, estrae ad intervalli regolari, un dato numero di informazioni chiamate *campioni*. Maggiore è la qualità dei campioni estratti, maggiore sarà la qualità finale ottenuta. Nel caso si prenda in considerazione un suono si parlerà di campionamento temporale, indicando con ciò ogni quanti secondi, o decimi, o millesimi di secondo le informazioni vengono estrapolate. Nel caso invece il campionamento riguardi un'immagine fissa, allora si parlerà di frequenza spaziale, indicando ogni quanti cm, mm, dm le informazioni vengono estrapolate. Quando il segnale è stato campionato bisogna utilizzare un quantizzatore che ha lo scopo di prendere i valori campionati e restituire un unico valore in uscita. Questo avviene poiché il range di valori di un campionamento può essere infinito mentre la codifica in digitale ha un range di valori finiti. Nel caso di una immagine a scala di grigio codificata in 8 bit ($2^8 = 256$), verrà assegnato ad ogni pixel un valore numerico da 0 a 255, espresso sotto forma binaria, indicante la gradazione dei

pixel. Nel caso di immagini più complesse e con una maggiore risoluzione, il singolo pixel potrà assumere un numero di stadi maggiori e un numero maggiore di bit necessari per rappresentarne in maniera univoca il colore.

1.5 Differenze tra analogico e digitale

I segnali analogici sono una trasmissione di uno spettro di informazioni che si diffonde per mezzo di un'onda. Il corpo umano, così come tutti gli esseri viventi, ha un corpo programmato per ricevere e interpretare queste onde analogiche. L'interpretazione di queste onde non traduce solo i punti massimi e minimi ma anche quelli intermedi. Ci rendiamo conto molto facilmente di quanto queste onde analogiche siano sensibili alle interferenze esterne. Se proviamo ad avere una conversazione in una stanza rumorosa oppure proviamo a sintonizzare la radio sulla giusta frequenza, notiamo che il suono è percepito in maniera distorta. A differenza dei segnali analogici quelli digitali non possono assumere le stesse sfumature di un'onda analogica. Essi assumono soltanto due stadi, acceso e spento. Nei sistemi analogici al contrario di quelli digitali non si converte nulla a 0 e 1 ma si utilizzano delle comparazioni reali. Nel caso di due orologi, uno analogico e uno digitale, la differenza sta nel fatto che l'orologio analogico mostra il tempo sulla comparazione spaziale mentre l'orologio digitale mostra il tempo direttamente sotto forma numerica. Bisogna però prestare particolare attenzione al fatto che non si tratta solo di una differenza visiva; la reale differenza tra i due sistemi si basa sui loro meccanismi e sul calcolo da loro utilizzato. Il meccanismo utilizzato dal sistema analogico prevede un sistema meccanico di ruote dentate che numera le oscillazioni di un bilanciere meccanico, nel caso del sistema digitale, invece, abbiamo un microprocessore che converte in bit le oscillazioni di un'onda elettrica all'interno di un circuito. Il digitale ha due grossi vantaggi rispetto all'analogico: meno interferenze e una maggiore risoluzione. La risoluzione dei segnali digitali è pressoché infinita poiché è possibile avere segnali digitali di qualità molto più elevata rispetto ai segnali analogici. Nel caso della registrazione di un suono, se questo viene registrato direttamente in digitale senza alcuna conversione, la precisione e la qualità del suono saranno superiori perché avremo un controllo maggiore e più facilmente gestibile del segnale.

1.6 Caratteristiche del documento digitale: vantaggi e problematiche

Quando si parla di digitale la prima cosa che viene messa in evidenza è il concetto di originale e di copia. In ambiente digitale questi due termini non hanno alcuna differenza in quanto un documento digitale è lo stesso indipendentemente dal numero di copie. Esso non è altro che una sequenza numerica binaria e per tale motivo è facilmente replicabile in un numero infinito di copie. I vantaggi della digitalizzazione sono legati alla trasformazione delle informazioni sotto forma di sequenza di bit, che può essere trattata in egual maniera da diversi supporti e da diversi computer. Inoltre, le trasformazioni in digitale sono prive, in generale, di errori mentre le informazioni analogiche introducono distorsioni e rumore. Per quanto semplice e immediata possa essere la copia di un documento digitale è al tempo stesso una operazione molto complessa e rischiosa. In primis dobbiamo ricordare che, a differenza di un documento analogico in cui figure simboli e frasi possono essere direttamente compresi (letti) da un essere umano, il contenuto di un documento invece è legato ad una tecnologia e non può essere direttamente interpretato da un essere umano. Dato che non è possibile leggere direttamente il formato digitale è molto importante che la sua decodifica segua esattamente le stesse specifiche adottate nella codifica. Avendo una rappresentazione numerica modificabile, il contenuto di un documento digitale gode della proprietà di non dover appartenere esplicitamente ad un dato supporto, ma può essere separato dal dispositivo originale e trasferito ad altri differenti tipi di strumenti di archiviazione. Se da un lato questo comporta un grosso vantaggio, dall'altro aumentano i rischi legati ad un'alterazione del documento stesso in relazione al mantenimento dell'autenticità e dell'attendibilità del documento. La capacità di replicare i documenti in maniera così semplice ci permette inoltre di farli sopravvivere nel tempo, ma la loro esistenza dipende non solo dalla facilità di duplicazione ma anche dal supporto in cui sono memorizzati. Essi infatti così come sono facilmente duplicabili sono anche soggetti ad un alto rischio di cancellazione per un errore da parte dell'utente o per un'alterazione dello stato magnetico di un settore del supporto di memorizzazione. La flessibilità del documento digitale permette di integrare in un unico documento qualsiasi tipo di materiale, trasformando il documento originale in un documento formato da varie

tecnologie e proprietà che ne amplificano le caratteristiche e lo rendono ancora più usabile. Grazie ad una serie di tecniche di controllo la trasmissione del documento digitale da pc a pc tramite supporto di memorizzazione o tramite la rete internet è resa ancora più sicura, ma anche in questo caso si corre il rischio di far intercettare⁵ il documento durante la fase di trasmissione.

1.6.1 Flessibilità

Una delle prime caratteristiche fondamentali dell'informazione digitale è la flessibilità. Un documento digitale come abbiamo detto, è una combinazione finita di valori o numeri 0 e 1. Più è grande il numero dei valori presenti all'interno del documento maggiore sarà la qualità e la rappresentazione reale dello stesso. Nel caso di un file audio abbiamo una rappresentazione di 5 miliardi di bit circa. Questa rappresentazione ci consente di ascoltare un brano musicale ad altissima fedeltà a patto di disporre di un dispositivo che consenta di riprodurre tale flusso numerico. L'approccio al digitale avviene tramite strumenti di conversione analogico/digitale poiché nessun messaggio si trasmette tale e quale in un ambiente ma subisce sempre delle trasformazioni. Una di queste è quella che unisce gli apparati cognitivi dell'uomo e del computer. Il nostro apparato uditivo non riesce a riconoscere il flusso numerico, abbiamo quindi bisogno di dispositivi in grado di convertire questo flusso di numeri in vibrazioni e viceversa. Poter gestire un flusso di numeri ci consente una versatilità fuori da ogni schema. Infatti è sufficiente invertire alcuni 0 e 1 per eliminare rumori di fondo, trasformare la tonalità della voce e degli strumenti, correggere distorsioni. Possiamo ad esempio racchiudere in un unico documento diversi tipi di informazioni, dalla scrittura, alle immagini, al suono. Sarà poi possibile avere innumerevoli copie dello stesso documento con un solo click del mouse e trasmetterlo in un qualsiasi dispositivo senza perdita di informazioni, tempo e costi.

1.6.2 Riproducibilità e conservazione

Entrambi i termini sono direttamente proporzionali l'uno all'altro in quanto la vita di un documento digitale è direttamente proporzionale alla sua facile riproducibilità. Vi

⁵Detto anche sniffing è l'attività di intercettazione passiva dei dati che transitano in una rete telematica

sono diversi rischi per un documento digitale di non sopravvivere nel tempo. Uno dei primi fattori di rischio è collegato direttamente al suo supporto di memorizzazione. Per quanto sia facile copiare un documento digitale e replicarlo, è altrettanto facile cancellarlo magari inavvertitamente con un semplice tasto della tastiera. Nel caso dei dischi magnetici invece i numeri che compongono il documento digitale possono essere alterati attraverso una fonte magnetica che ne altera la sequenza. Come abbiamo la possibilità di manipolare con un solo click la combinazione numerica, allo stesso modo è possibile con un solo click cancellarla ripristinando lo stato originale “*di tabula rasa*”. Anche per quanto riguarda la memorizzazione su dispositivi ottici (CD/ DVD) la possibilità di salvare tali documenti in supporti che non è possibile modificare o cancellare dopo la loro scrittura, esclude i vantaggi connessi alla flessibilità del documento digitale. Su opere come libri, tele e documenti cartacei abbiamo informazioni ben precise sulla durata degli stessi nel tempo, ma per quanto riguarda i sistemi tecnologici come i cd o gli hard disk, le nostre informazioni scarseggiano. Tali supporti esistono da troppo poco tempo per poter dare una stima e fornire delle previsioni. Tra i fattori che influenzano maggiormente la conservazione e la riproducibilità di un documento digitale vi sono senza dubbio gli strumenti di lettura dei supporti. Solo un refreshing⁶ (copiare) dai vecchi ai nuovi supporti garantisce la possibilità di riproduzione e conservazione. Vi sono infatti numerose e differenti tecnologie tra loro incompatibili che non permettono una lettura delle informazioni in maniera universale.

1.6.3 Trasmissibilità

La trasmissione delle informazioni dal mittente al destinatario spesso non ammette errori. Se si verificano delle interferenze il destinatario otterrà un messaggio incompleto e distorto. Nelle trasmissioni analogiche l'unica possibilità per ovviare a questi inconvenienti è lasciata all'interlocutore finale e alla sua capacità di interpretazione o alla ritrasmissione del segnale. Nel caso di una trasmissione digitale dall'utente iniziale all'utente finale, la trasmissione di bit deve pervenire all'utente finale nella medesima sequenza in cui è stata spedita indipendentemente dal tipo di supporto in cui è stata

⁶È il trasferimento di dati tra due supporti digitali dello stesso tipo; in questo modo non ci sono cambi o alterazioni del bitrate dei dati

trasmessa o registrata. Durante la trasmissione la ricorversione del messaggio digitale è rappresentata sotto forma di grandezza fisica e più precisamente da una sinusoide.

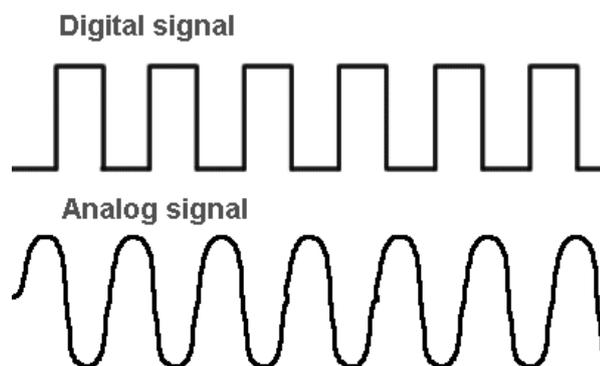


Figura 1.1: Segnale digitale vs analogico

Essa rappresenta un'onda acustica. Dato che tutte le informazioni trasmesse in modo analogico hanno una loro onda e una loro ampiezza e dato che nessuna onda analogica durante la trasmissione è priva di mutamenti di ampiezza, la riconversione dall'analogico al digitale traduce in 1 tutto ciò che sta sopra lo 0 e in 0 tutto ciò che sta sotto lo 0. È ovvio che una tale ricampionatura elimina tutte le sfumature dell'onda ed è per questo che bisogna operare con una adeguata frequenza. È possibile inoltre attuare delle strategie automatiche di controllo che assicurano che la ricezione dell'informazione digitale giunga al destinatario nell'ordine esatto in cui è stata inviata, come nel caso del protocollo di comunicazione TCP⁷ (Transfert Control Protocol), uno dei protocolli alla base del funzionamento di internet. Nel momento in cui è presente un errore di trasmissione e il ricevente non ha ottenuto tutte le informazioni, il protocollo TCP fa scattare automaticamente una procedura di correzione che consiste nella richiesta di una seconda spedizione da parte del mittente dei soli pacchetti mancanti. Inoltre la trasmissione del messaggio digitale può essere resa meno pesante attraverso una sua ricampionatura⁸ chiamata "*comprensione*" in cui attraverso un algoritmo si trasforma la

⁷È un protocollo di rete a pacchetto di livello di trasporto, appartenente alla suite di protocolli Internet, che si occupa di controllo di trasmissione

⁸<http://pcfarina.eng.unipr.it/dispense01/onofri132321/onofri132321.htm>

sequenza numerica in espressioni che saranno poi riconvertite dall'utente finale, procedura che fa risparmiare tempo e costi di spedizione.

Capitolo 2

Gestione dei documenti in ambiente digitale

La conservazione dei dati digitali è una procedura nata da poco tempo. I problemi di archiviazione e conservazione non dipendono solo dall'utilizzo del supporto hardware e software e dalla loro relativa obsolescenza ma anche, dall'utilizzo di metodi di conservazione condivisi, univoci e regolamentati. La fragilità dei supporti è una parte marginale del problema. L'evoluzione continua della tecnologia permette di intervenire sui dati per correggere errori e trasferirli da supporti di vecchia generazione a supporti di nuova generazione in maniera sempre più sicura. La conservazione invece manca di standard internazionali, norme nazionali, procedure interne. Ciò che viene affrontato qui di seguito è l'analisi di due problemi:

- Rendere i contenuti accessibili a chiunque;
- Rendere il testo accessibile nel tempo.

2.1 La conservazione

Le soluzioni tecnologiche e organizzative insieme ad una analisi adeguata devono consentire:

- memorizzazione e 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.

Per poter usufruire dell'accesso ai dati in maniera integra e ridurre allo stesso tempo i costi per la manutenzione, è importante intervenire a monte durante la fase di creazione. Tutti i dati che nascono in un ambiente non definito e non controllato, corrono un alto rischio di venir danneggiati o persi. La conservazione dei dati digitali è definita tale solo quando è a lungo termine, priva di eventuali errori di conservazione ed è accompagnata dall'inclusione di metadati di preservazione. Prima di essere un fattore tecnologico, deve essere considerato un fattore di natura socio-culturale, con forti implicazioni organizzative, economiche, giuridiche. L'eredità culturale digitale è una parte fondamentale della creatività e delle conoscenze dell'umanità che non può passare inosservata. Questo infatti va oltre il processo tecnico di conservazione. È un processo che implica la scelta di quali risorse conservare e in che forma.

2.2 Standard

Essendo un campo nato relativamente da poco tempo, è difficile poter trovare uno standard consolidato per gestire i molteplici aspetti della conservazione digitale. Sono stati fatti diversi progressi a tal proposito soprattutto per la definizione di modelli di archivi digitali e la definizione di metadati per la conservazione, ma ancora c'è molto da fare. L'Istituto Europeo per gli Standard nelle Telecomunicazioni (ETSI), è un organismo internazionale indipendente e senza fini di lucro responsabile della definizione e dell'emissione di standard nel campo delle telecomunicazioni in Europa. Nel 2011 sono stati rilasciati due importanti documenti specifici per il mondo digitale. I due documenti, ovvero il 101 533-1 e 101 533-2, concentrano la loro attenzione sulla sicurezza informatica che deve essere garantita al sistema di conservazione dei dati digitali. In particolare lo standard 101 533-1 individua i requisiti di sicurezza informatica che devono essere rispettati da un fornitore di servizi di conservazione, mentre lo standard 101 533-2 individua

le linee di requisiti da seguire per chi valuta il sistema informatico di conservazione. Ciò che i due standard mettono in evidenza non è il tipo di supporto da utilizzare, uno specifico linguaggio di programmazione, un formato o la struttura per realizzare cartelle virtuali o metadati, ma concentra la sua attenzione sugli aspetti strettamente legati alla sicurezza delle informazioni. Un aspetto rilevante viene dato alla gestione della sicurezza degli accessi al sistema, del disaster recovery¹, della indicizzazione dei dati. Entrambi i documenti si rifanno allo standard principale ETSI TS 102 573 “Policy requirements for trust service providers signing and/or storing data for digital accounting”. Una particolarità interessante di tali documenti è l’accento posto sull’utilizzo degli Open Source² come sistema che garantirebbe interoperabilità e l’indipendenza dall’obsolescenza tecnologica di tutti i sistemi presenti sul mercato. Sul fronte italiano l’UNI (Ente Nazionale Italiano di Unificazione), associazione privata senza scopi di lucro, rappresenta l’Italia nelle attività di normazione. È formato da numerosi sotto enti tra i quali UNINFO, che rappresenta la normalizzazione delle attività in ambito informatico a livello nazionale ed internazionale presso l’ISO, l’ISO/IEC JTC 1 e il CEN. L’ISO (Organizzazione Internazionale per la Normazione) è la più importante organizzazione a livello mondiale per la definizione di norme tecniche. Molti dei supporti di memorizzazione dati come il CD o il DVD utilizzano l’estensione ISO, ad indicare che tali supporti utilizzano il file system ISO 96690.

2.3 Metadati

Quando si converte un dato da analogico a digitale non si codifica solo l’informazione in se ma il dato digitale viene incorporato all’interno di una struttura di metadati. Letteralmente sono definiti come “dati sui dati”, informazioni strutturate e suddivise per campi, che ne garantiscono una maggiore organizzazione e recupero. L’utilizzo di questi metadati ha un impatto notevole nel mondo digitale, soprattutto nel cyberspazio per

¹Si intende l’insieme di misure tecnologiche e organizzative atte a ripristinare sistemi, dati e infrastrutture necessarie all’erogazione di servizi di business a fronte di gravi emergenze

²http://it.wikipedia.org/wiki/Open_Source_Definition

ridurre al minimo l'information overload³. È possibile eseguire numerose informazioni quali ad esempio autenticazione ed autorizzazione, personalizzazione e localizzazione dei servizi, gestione delle informazioni, protezione dell'autenticità dei dati. Dato l'alto numero di operazioni che è possibile eseguire, non è possibile trovare un'unica soluzione per tutti gli usi. È necessario perciò specificare i requisiti funzionali per definire le caratteristiche dei metadati da utilizzare. Oltre alla conservazione del flusso di bit, i metadati servono a mantenere le informazioni necessarie a riconoscere i dati digitali nei futuri supporti di memorizzazione. Sono fondamentali per mantenere i dati accessibili, utilizzabili e autentici (identificabili ed integri) nel medio e lungo periodo. Maggiore sarà la completezza dei metadati, minore sarà l'intervento di recovery sui dati digitali in futuro. È necessario tener traccia dei metadati utilizzati, della loro relazione con i dati attraverso una accurata documentazione. È possibile suddividere i metadati in 3 sottocategorie:

- Descrittivi: sono orientati al recupero degli oggetti digitali;
- Amministrativi e gestionali: per le operazioni di gestione degli oggetti digitali all'interno degli archivi;
- Strutturali: descrivono la struttura fisica interna o logica dei documenti.

I metadati amministrativi e gestionali offrono un supporto maggiore per le maggiori informazioni sulla creazione, immissione e manutenzione dei dati digitali. Inoltre i metadati devono essere considerati come dinamici poiché, da quando nascono fino alla durata di tutto il ciclo di vita di una risorsa o dato, sono in continua evoluzione. In principio i metadati furono utilizzati da professionisti ed in particolare da archivisti e bibliotecari per poter catalogare e descrivere i documenti. Con l'avvento del cyberspazio⁴ anche gli utenti meno esperti sono stati e sono potenzialmente in grado di creare metadati che descrivono i documenti da loro stessi creati. Oggi questi dati su dati possono essere direttamente creati in maniera automatica dal computer con un alto livello di complessità di

³Sovraccarico informativo è un termine inglese usato per indicare una quantità di informazione eccessiva senza che vi sia un reale motivo della sua esistenza

⁴È dimensione astratta che permette ai computer di tutto il mondo di comunicare in un'unica rete ed agli utenti di relazionarsi tra loro

relazioni tra diversi computer interconnessi. In generale i metadati, indipendentemente dagli oggetti informativi, hanno tre caratteristiche che li accomunano:

- Il contenuto che stabilisce un nesso tra il dato e ciò che lo caratterizza e lo circonda;
- Il contesto che indica il chi, cosa, perché, come, legati alla creazione dell'oggetto;
- La struttura, insieme formale di associazioni situate tra gli oggetti informativi all'interno del oggetto.

Si prefiggono di garantire diversi obiettivi come l'identificazione permanente dei singoli oggetti informativi, l'identificazione permanente delle relazioni tra gli oggetti informativi, conservare le informazioni che supportano l'intelligibilità degli oggetti informativi, mantenere le informazioni relative alle attività di tenuta rilevanti ai fini dell'autenticità.

2.4 I codici in uso

Il numero binario che viene assegnato al dato codificato varia a seconda del codice di codifica usato. Vengono usati vari sistemi di codifica tra cui ASCII, EBCDIC, UNICODE.

- ASCII (American Standard Code for Information Interchange) è stato sviluppato dall'American National Standard Institute (ANSI) per la codifica di una serie di caratteri a 7 bit. Si differenzia dalla versione a 8 bit con cui ci si riferisce con la sigla US-ASCII. È stato proposto e realizzato nel 1961 da un ingegnere dell'IBM Bob Bemer, successivamente accettato come standard ISO (ISO 646). La ragione per cui è stata introdotta la versione a 8 bit deriva da fatto che, 7 bit non sono sufficienti a rappresentare la maggior parte dei caratteri. La versione a 8 bit permette di raddoppiare il set di caratteri rappresentabili. L'ASCII sta per lasciare il posto al suo successore UTF-8, ormai riconosciuto ed usato da ogni programma in circolazione. UTF-8 sfrutta la potenzialità creata dall'ASCII sul bit di controllo⁵,

⁵È un codice di controllo utilizzato nei calcolatori per prevenire errori nella trasmissione o nella memorizzazione dei dati

usandolo non più come bit di controllo ma come indicatore. Se al posto del vecchio bit c'è 0 allora sarà letto come ASCII a 7 bit, altrimenti se c'è 1 viene concatenato al byte successivo. In tal modo si riescono ad includere tutti gli alfabeti e potenzialmente può essere ancora esteso per rappresentarne altri;

- EBCDIC (Extended Binary Coded⁶ Decimal Interchange Code) è la codifica più usata nei mainframe e nei server più potenti. Ha una codifica a 8 bit ed è stata creata tra il 1963 e il 1964. Inizialmente fu usato per la sua facilità di utilizzo nei sistemi che utilizzavano le schede perforate. Poiché tale supporto è ormai del tutto obsoleto, negli elaboratori moderni il sistema viene usato soltanto per una compatibilità dei sistemi più vecchi. Infatti non è presente alcun vantaggio tecnico rispetto alla codifica ASCII. Questa codifica non consente l'utilizzo di più lingue in maniera contemporanea;
- UNICODE dotato di oltre 96 mila caratteri, è in grado di rappresentare le principali lingue nel mondo. Viene assegnato un numero univoco ad ogni carattere utilizzato per la scrittura di testi in maniera indipendente dalla lingua, dalla piattaforma e dal programma utilizzati. Attualmente ancora non rappresenta tutti i caratteri del mondo essendo in evoluzione ma si sta prefiggendo di raggiungere presto questo scopo garantendo la massima compatibilità e la non sovrapposizione con altre codifiche già in uso.

2.5 Formati e metalinguaggio

Il metalinguaggio è un linguaggio formato da regole sintattiche, usato per descrivere altri linguaggi. Si tratta di uno strumento che ha delle precise forme sintattiche da rispettare scritte in modo non ambiguo. I linguaggi di markup descrivono i meccanismi di rappresentazione di un dato che può essere usato su più supporti. Sono i linguaggi più opportuni per strutturare e marcare i dati digitali in maniera indipendente dall'applicazione, favorendo la riusabilità e la flessibilità. I markup in circolazione sono stati

⁶È un programma o un dispositivo che si occupa di codificare e/o decodificare digitalmente un segnale (tipicamente audio o video) perché possa essere salvato su un supporto di memorizzazione o richiamato per la sua lettura

creati sia da aziende private, proprietari, sia da individui o aziende non commerciali, non proprietari. Per i primi l'azienda detiene i diritti e può modificarli, aggiornarli e rivoluzionarli in qualunque momento e per qualunque motivo. Il secondo invece è stato creato da un gruppo di individui o aziende che non sono commerciali e che seguono un interesse comune che risponde alle esigenze di ciascun partecipante. Esistono diversi metalinguaggi che possono essere elencati come segue:

- DublinCore nato da un gruppo organizzato dall'Online Computer Library Center (OCLC), che ha sede a Dublin, in Ohio (USA), dal quale prende il nome, e dal National Center for Supercomputer Application (NCSA). Lo scopo era di stabilire un insieme base di elementi per sviluppare un metalinguaggio per la descrizione delle informazioni elettroniche in rete. Ha una forma molto semplice per permettere anche a chi non fosse uno specialista, di poter creare agevolmente ed economicamente semplici record di descrizione. Ma questa sua semplicità diventa un limite quando bisogna descrivere risorse specifiche in modo adeguato. Questo costringe a utilizzare lo schema e a personalizzarlo per soddisfare le esigenze particolari senza garantirne l'interoperabilità;
- RDF (Resource Description Framework) è stato proposto dal W3C per la codifica, lo scambio e il riutilizzo di metadati mantenendo l'interoperabilità tra le applicazioni che si scambiano le informazioni sul WEB. Tutte le informazioni descritte da RDF sono definite come risorse intese non solo come risorse informatiche ma anche come risorse che non risiedono sul web. Ogni risorsa è identificata da un URI⁷ (Universal Resource Identifier), proprietà e valori. Le proprietà individuano le relazioni tra le risorse e i valori e sono a loro volta identificate da un URI, mentre il valore è un tipo di dato primitivo;
- Marc (Machine-Readable Cataloging) fu ideato intorno agli anni 60 per rispondere all'esigenza di ridurre al minimo lo spazio occupato per l'archiviazione e trasmissione delle informazioni digitali senza però rendere troppo complessa la gestione.

⁷È una stringa che identifica univocamente un indirizzo WEB, un documento, un'immagine, un file, un indirizzo di posta elettronica, ecc, che è presente sul WEB http://it.wikipedia.org/wiki/Uniform_Resource_Identifier

Quello che il linguaggio cercava di ottenere era una definizione semantica e sintattica della rappresentazione dell'informazione. A lungo andare però questo obiettivo ha fatto nascere grossi problemi di gestione dell'informazione. Fu deciso quindi di separare la parte sintattica dalla semantica, definendo uno standard più rigido e generalizzato. In seguito nacquero numerosi altri metalinguaggi come MARC21, USMARC, FINMARC, DENMARC, UKMARC;

- Z39.50 è un protocollo di comunicazione client-server creato da ANSI/ISO (American Institute, National Information Standards Organization) ed ISO 23950 per organizzare le informazioni presenti in nel WEB e renderle più facilmente accessibili. Particolarmente usato per consultare le basi di dati di qualsiasi biblioteca indipendentemente dalla piattaforma utilizzata poiché utilizza un unico linguaggio di interrogazione e un'unica interfaccia. Altra importante caratteristica sta nella sua capacità di trasmettere contemporaneamente la richiesta a più database. Vi sono ancora dei problemi relativi alla risposta dell'interrogazione poiché il risultato spesso torna risultati doppi. Dalla sua nascita ha subito delle modifiche e delle evoluzioni che hanno portato questo protocollo a raggiungere la terza versione.

Altri metalinguaggi hanno visto crescere il loro uso per via di caratteristiche e strutture che nel tempo sono state condivise dalla maggior parte degli utenti e che hanno ancora oggi un ruolo centrale. SGML, XML, HTML sono metalinguaggi molto più comuni che sono prevalsi con maggiore enfasi su altri tipi di metalinguaggi per la loro flessibilità, struttura, usabilità.

2.5.1 SGML

Standard Generalized Markup Language (SGML) è un metalinguaggio non proprietario di markup descrittivo definito dallo standard ISO 8879:1986 SGML e realizzato da Charles Gouffard per la creazione e lo scambio di documenti elettronici. Essendo un metalinguaggio non proprietario garantisce che il controllo non sia detenuto da un'unica casa produttrice e ciò assicura una continua evoluzione e una accettazione generale delle sue regole e del suo utilizzo. Non è un linguaggio di markup ma è un linguaggio con cui

si definiscono linguaggi di markup. Il suo uso è destinato per tutti i documenti analogici che devono essere codificati, trasmessi e archiviati con strumenti informatici. Non dipende da un unico programma o da una singola architettura e la portabilità dei dati è garantita su qualsiasi tipo di piattaforma senza perdita di informazioni. Il markup è posizionato accanto agli elementi del testo a cui si riferisce e può essere letto indistintamente da un essere umano o da un programma. La struttura è di tipo gerarchica e la sua funzione principale è il DTD⁸ (Document Type Definitio) che definisce in maniera rigorosa la struttura logica che devono possedere questi tipi di documenti. Tale struttura logica è formata da elementi come nome, opera, autore, collocazione e le relazioni che intercorrono tra loro. Inizialmente usato per progetti governativi e militari, il markup aveva il compito di assicurare la leggibilità del documento per diversi anni. L'SGML è molto flessibile, potente e i vantaggi del suo utilizzo sono molteplici:

- è indipendente da qualsiasi tipo di architettura e questo consente di leggerlo su ogni tipo di computer e di sistema operativo;
- può essere utilizzato da diverse applicazioni, proprietarie e non, senza che ci sia una codifica da un formato all'altro;
- anche se strettamente nato per documenti elettronici questo standard può essere utilizzato su un qualsiasi tipo di dato digitale;
- ha una struttura espandibile e ben strutturata che consente di essere sfruttata per risolvere problemi di elaborazione dell'informazione molto complessi, garantendo la massima usabilità dei dati.

In particolar modo questo metalinguaggio è stato usato nel settore umanistico con una applicazione definita con il nome di TEI (Text Encoding Initiative) utilizzata come standard per la codifica e l'archiviazione digitale dei libri per la creazione di biblioteche digitali.

⁸È uno strumento utilizzato dai programmatori il cui scopo è quello di definire le componenti ammesse nella costruzione di un documento XML

2.5.2 XML

eXtensible Markup Language (XML). Definisce e controlla il significato degli elementi contenuti in un documento o in un testo. È un tentativo di semplificare la versione SGML per rappresentare in maniera semplice nuovi linguaggi di markup utilizzati nel web, in particolare ha semplificato la struttura eliminando le funzioni più complesse e rendendo facoltativo l'utilizzo del DTD. Inoltre una caratteristica molto importante del linguaggio XML è data dalla possibilità di utilizzare link ipertestuali che nel linguaggio SGML non sono stati implementati. È definito dal consorzio World Wide Web Consortium (W3C), l'organismo che definisce tutti gli standard utilizzati nel WEB. Tutte le informazioni relative al linguaggio e alle sue regole sono depositate sul sito W3C. Nel caso dell'XML si tratta di un linguaggio marcatore (markup language) estendibile (eXtensible) che consente di creare tag⁹ personalizzati. La sintassi utilizzata da XML è molto rigida e per tale ragione bisogna rispettare alcune regole altrimenti non verrà correttamente implementato dal browser¹⁰. Se è ben strutturato un file XML è detto Well Formed (scritto correttamente), e se in più rispetta i requisiti definiti nel file DTD allora sarà chiamato Valid (Valido).

2.5.3 HTML

HyperText Markup Language (HTML) è un linguaggio molto comune usato sul web per creare documenti ipertestuali. È caratterizzato da una struttura che fa uso di tag che definiscono le proprietà che il documento deve assumere, dal colore alla grandezza del testo, al link. In generale quando si accede sul web il client fa una richiesta al server che ne restituisce un documento HTML e file ad esso correlati. Nonostante la sua struttura non bisogna confonderlo come linguaggio di programmazione ma come semplicemente un linguaggio di markup. Supporta l'inserimento di script e oggetti esterni. La prima versione ufficiale fu rilasciata dal W3C nel 1997 denominandola HTML 4.0. È un linguaggio chiuso e non è possibile modificarne la struttura. È composto da tag ben strutturati che inglobano le informazioni ma che ne impediscono di strutturarle secondo livelli gerarchici. Rispetto agli altri linguaggi per il WEB come ASP, PHP, è in grado di

⁹Termine associato a un contenuto digitale per facilitarne l'indicizzazione nei motori di ricerca

¹⁰<http://it.wikipedia.org/wiki/Browser>

gestire i documenti in modo dinamico. Con l'ultima release, HTML5, questo linguaggio ha raggiunto un notevole traguardo. Sono stati aggiunti nuovi comandi e funzionalità fino a quel momento ottenute tramite una procedura non standard e il raggiungimento di una buona compatibilità con i browser esistenti, compresi quelli presenti su piattaforma mobile o obsoleti.

2.5.4 OAIS

Open Archival Information System (OAIS) è lo standard ISO:14721:2003 di conservazione dei dati digitali a lungo termine. Non è un modello che preserva esclusivamente i bit ma detta regole e direttive per rendere usabile un qualsiasi oggetto digitale. In particolare descrive l'ambiente in cui il dato digitale risiede, i componenti funzionali dell'archivio stesso, l'infrastruttura delle informazioni che supporta i processi dell'archivio. Questo modello ha una struttura logica e non dipende da nessuna tecnologia in circolazione. Per tale motivo è un modello che viene preso sempre più in considerazione. Ciò che descrive il modello è una relazione tra uomo e sistema che operano insieme per conservare determinate informazioni da mettere a disposizione di una cerchia definita di utenti. Risulta subito chiaro il processo di elaborazione dello standard realizzato attraverso il contributo di tutti i soggetti interessati. Il modello rappresenta un luogo di confronto e valutazioni di soluzioni differenti adatto a qualsiasi tipo di sistema, sia esso archivio, biblioteca o altro ente. Stabilire i criteri di selezione e acquisizione, ottenere un controllo delle risorse archiviate, determinare la comunità di riferimento, assicurare che l'informazione archiviata sia comprensibile in modo autonomo dalla comunità degli utenti senza l'intervento dei soggetti che hanno prodotto l'informazione, adottare politiche e procedure che garantiscano garanzia di sicurezza in materia di conservazione, rendere disponibile all'utente le risorse dell'archivio, sono punti guida ben definiti all'interno del modello.

Capitolo 3

Le memorie di massa

3.1 Classificazione delle memorie

La memorizzazione dei dati digitali può sfruttare ad oggi una tipologia di memorie di massa molto varia. Esse possono essere catalogate sotto differenti aspetti quali gerarchia, tipo di accesso, permanenza nei supporti, tecnologia impiegata. Una prima classificazione in base alla gerarchia racchiude le cosiddette memorie primarie e secondarie:

- La memoria primaria (detta anche memoria principale o centrale) è una memoria costituita dalla RAM, dalla ROM e dalla memoria Cache;
- La memoria secondaria (detta anche memoria esterna o di massa), è costituita da memorie in grado di memorizzare grosse quantità di dati come nastri magnetici, dischi fissi, CD, DVD, memorie flash;

Una seconda classificazione può essere fatta in base al tipo di accesso al dato. L'accesso può avvenire tramite:

- accesso casuale (o diretto), riguarda il tempo necessario per accedere al dato che non dipende dalla posizione fisica (settore) del dato sulla memoria estesa (RAM, ROM, dischi fissi e ottici);
- accesso sequenziale, quando il tempo di accesso al dato dipende invece dalla posizione in cui il dato si trova sulla memoria di massa (nastri magnetici, dischi ottici).

Una terza classificazione riguarda la persistenza dei dati nelle memorie ed è costituita da:

- memorie temporanee (o volatili) quando la perdita dei dati avviene istantaneamente al venir meno della corrente elettrica (RAM);
- memoria permanente (o non volatile) quando i dati non vengono persi in assenza di corrente elettrica (dischi fissi, memorie flash, supporti ottici).

Una quarta classificazione riguarda la possibilità di scrittura da parte dell'utente:

- le memorie Read Only possono essere solo lette, come nel caso della ROM (Read Only Memory);
- le memorie WORM (Write Once, Read Many) possono essere scritte solo una volta e lette più volte, come i CD o i DVD;
- le memorie WMRM (Write Many, Read Many) possono essere scritte più volte e lette più volte, come nel caso dei dischi fissi, CD-RW e i DVD+/- RW.

Una quinta ed ultima classificazione riguarda la tecnologia impiegata:

- memoria cartacea;
- memorie magnetiche, che utilizzano campi elettromagnetici (floppy disk e hard disk).
- memorie ottiche, che utilizzano un raggio laser per leggere e scrivere i supporti.
- memorie elettroniche, che si basano su componenti elettronici come i transistor.
- memorie olografiche, che si basano su tecnologia omonima.

3.2 Tipi di memorie di archiviazione

3.2.1 Memoria cartacea

Per quanto inverosimile e molto lontana dalla moderna tecnologia di archiviazione ha rappresentato la prima memoria di massa. Si trattava di schede perforate (punch card) e dei nastri perforati (punch tape) utilizzati tra gli anni '50 e '70. Il loro funzionamento si basava su rulli di carta perforati sui quali venivano incisi dei fori che rappresentavano la sequenza di bit 0 e 1. Erano estremamente lenti e scomodi da usare e presto furono sostituiti dall'avvento dei nastri magnetici.

3.2.2 Memoria magnetica

È stata la prima forma di memoria di massa. L'idea che sta alla base di questo tipo di memorizzazione è basata sul principio di funzionamento del magnetismo. In particolare si parla di dati memorizzati sotto forma di stati magnetizzati della materia. In un dispositivo a memoria magnetica si utilizza l'azione di attrazione e respinta di due poli magnetici e attraverso una opportuna emissione di corrente elettrica, la testina è in grado di allineare le cariche magnetiche presenti sulle superfici consentendo la scrittura delle informazioni. La superficie dei dischi viene suddivisa virtualmente in diverse aree o settori in cui la testina può leggere o scrivere informazioni. I settori si alternano sotto la testina man mano che il disco gira. Ciascun settore e ciascuna traccia vengono a loro volta suddivisi in singoli blocchi di dati che conterranno al loro interno un determinato numero di semplici informazioni. I dati possono essere scritti, cancellati e riscritti per un numero indefinito di volte senza che il supporto subisca danni. Ciò che ne influenza il funzionamento sono i campi elettromagnetici che rendono altamente vulnerabile il supporto e per tale motivo è fortemente consigliabile tenere questi supporti lontano da sorgenti elettromagnetiche come trasformatori di potenza o grosse calamite.

Floppy disk Sono dischi rimovibili che fino a poco tempo fa sono stati i più comuni dischi di archiviazione per il trasferimento di dati da un pc ad un altro per via della loro facilità di trasporto e del loro costo. La loro capacità di memorizzazione è molto bassa e questo è stato uno dei fattori principali della loro sostituzione nel tempo. Il

disco è contenuto in un supporto di plastica molto leggero e flessibile. Per poterli utilizzare bisogna possedere l'apposito drive¹. Una volta polarizzate, le particelle sul disco non si smagnetizzano spontaneamente quindi l'informazione resta tale fino ad un nuovo cambio di polarizzazione della particella. Il disco è forato al centro e ruota sul suo asse. Una testina di lettura/scrittura si muove per mezzo di un braccio meccanico che la sostiene e per mezzo di scatti, si posiziona sulle tracce. Per poter utilizzare questo supporto bisogna inicializzarlo, o per meglio dire formattarlo. La formattazione del supporto consiste nel suddividere il disco in settori aventi la stessa ampiezza angolare. I blocchi in cui vengono memorizzate le informazioni risiedono tra i vari settori delimitati dalle tracce disposte in maniera concentrica e che intersecano i settori. I blocchi ottenuti da queste suddivisioni possono contenere un indirizzo con qualche centinaia di byte di memorizzazione. I dati memorizzati possono occupare diversi blocchi ed è per questo che ogni blocco di inizio scrittura dei dati è inicializzato con un indice che prende il nome di FAT² (File Allocation Table), ed indica il blocco fisico nel quale il dato memorizzato. La dimensione dei floppy è passata dagli 8 pollici per il primo floppy disk uscito nel 1971 con 360 kb di capacità, ai 5 pollici del 1976 con 720 kb, ai 3 pollici e mezzo del 1981 con 2,88 MB. Essi sono stati prodotti in quattro formati: SD (Single Density), DD (Double Density), HD (High Density), ED (Extended Density). Simili ai floppy disk sono stati creati e utilizzati i cosiddetti dischi Zip, dischi LS/120, dischi Syquest. Cioè che fondamentalemente li differisce è la loro capacità di memorizzazione molto più elevata: dai 100 ai 700 MB per i Zip, da 1 o 2 GB per i dischi LS/1002, da 230 MB ai 1,5 GB per i Syquest.

Nastri magnetici Appartengono alle memorie ad accesso sequenziale usate nei grandi sistemi informatici per creare copie di backup di dati importanti. Questi supporti sono costituiti da un materiale plastico che ricopre decine di metri di un nastro rivestito di materiale magnetizzabile. Il loro meccanismo di memorizzazione è lo stesso dei floppy disk. Le cartucce utilizzano un apposito drive chiamato tape

¹Permette al sistema operativo di utilizzare l'hardware senza sapere come esso funzioni, ma dialogandoci attraverso un'interfaccia standard

²http://it.wikipedia.org/wiki/File_Allocation_Table



Figura 3.1: Floppy-Disk

streamer che è in grado di memorizzare alcune centinaia di GB. In questo caso però la fase di lettura e scrittura è molto lenta e per tale motivo il loro utilizzo viene impiegato nelle ore notturne. I file sono memorizzati in sequenza e questo costringe a leggere tutti i dati precedenti prima di leggere quello desiderato. La difficoltà di questi supporti è data dalla poca flessibilità nel memorizzare i dati in quanto uno stesso file precedentemente memorizzato con una capienza minore, in previsione di un aumento di dimensione, costringe i computer a memorizzarlo in fondo al nastro portando all'inutilizzazione di parte del nastro. Al contrario nel caso di riduzione della dimensione del file andrebbe sprecata la parte del nastro già occupata. Anche in questo caso sono stati creati diversi supporti: DAT (Digital Audio Tape), DLT (Digital Linear Tape), LTO (Linear Tape Open), AIT (Advanced Intelligent Tape). La vita di questi supporti si aggira tra i 10-30 anni.

Disco rigido È la memoria di massa principale utilizzata nei computer. È utilizzato per memorizzare tutti i dati, compreso il sistema operativo, e può essere considerato come una estensione del floppy disk. Come nei floppy disk anche l'hard disk possiede un disco suddiviso in settori, blocchi, tracce, testina e braccio meccanico, ma in questo caso non vi è soltanto un disco ma una serie di dischi che ruotano su un asse comune e sono letti e scritti da un gruppo di testine. I dischi sono formati da vario materiale, in genere vetro e alluminio ricoperti da particelle ferromagnetiche che girano ad una velocità di 5200, 5400, 7200 e 10000 rpm (giri al minuto). Le testine di

lettura e scrittura hanno dimensioni molto piccole e non hanno alcun contatto fisico con la superficie del disco perché tra la testina e il piatto si crea un cuscinio d'aria che viene generato per via dell'alta velocità di rotazione dei dischi. Quando l'hard disk è spento la testina viene posizionata in una parte denominata landing zone, per evitare un possibile contatto con la superficie dovuto a cause accidentali come urti o cadute. Per ogni disco ci sono due testine che vengono mosse da un attuatore elettromeccanico. Avente una struttura meccanica più delicata rispetto ai floppy disk, tutto il gruppo formato dalle testine e dai dischi è collocato all'interno di un contenitore sigillato che li protegge dagli agenti esterni. Anche per gli hard disk bisogna inizializzare i blocchi. Data l'elevata capacità di memorizzazione e la diversa grandezza fisica dei file che possono essere memorizzati, vengono utilizzati tipi diversi di formattazione che permettono di avere una più o meno elevata capacità di gestione dei file (FAT, FAT32, NTFS³, JOURNALING⁴, EXT4⁵). All'interno dei dischi l'informazione è memorizzata sul blocco. Se l'informazione aumenta la sua grandezza vengono accodati al blocco iniziale tanti blocchi quanti necessari a contenere le successive informazioni. Se vi è una diminuzione dell'informazione, i blocchi che erano stati accodati vengono liberati e messi nuovamente a disposizione per la memorizzazione di altri file. Normalmente è collocato internamente al computer e non è visibile né estraibile, da qui la definizione di disco fisso. Esistono anche dischi rigidi collegati esternamente (dischi rigidi portatili) da collegare via USB. I primi modelli avevano una scarsa capacità di memorizzazione dei dati ma con il tempo la capacità è aumentata passando così da pochi MB all'ordine di centinaia di GB o del TB. Vi è poi una interfaccia verso il computer per la trasmissione e la ricezione dei dati (ATA, EIDE, SCSI, SAS, SATA).

³<http://it.wikipedia.org/wiki/NTFS>

⁴È una tecnologia utilizzata da molti file system moderni per preservare l'integrità dei dati da eventuali cadute di tensione

⁵È un file system journaled sviluppato come successore di ext3, incluso nel kernel Linux

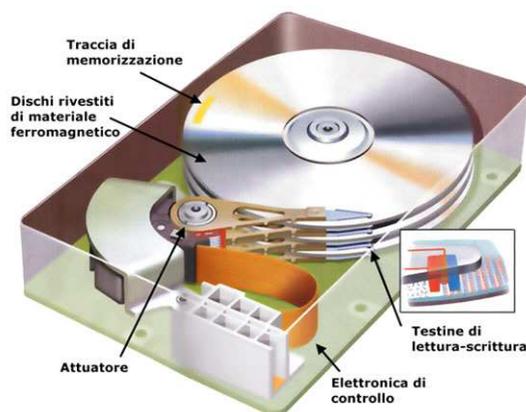


Figura 3.2: Hard-Disk

3.2.3 Memorie ottiche

Sono ampiamente usate per dati come audio e video, e per backup in quanto la loro capacità di memorizzazione è molto ampia. Viene utilizzata una forma di incisione dei supporti fisici, generalmente composti da polycarbonato trasparente in grado di riflettere la luce, nei quali viene inserito internamente un sottile foglio metallico, quasi sempre di alluminio. L'incisione avviene tramite un raggio laser al di sopra del quale sono montate una serie di lenti. Il laser viene proiettato sul supporto ottico che marca una serie di incisioni che identificano la sequenza binaria 0 e 1. Questo tipo di supporti memorizzano le informazioni su di un'unica traccia a spirale che parte dalla parte più interna del supporto fino alla parte più esterna. La lettura dei supporti avviene in maniera sequenziale seguendo l'aspirale. I supporti ottici di questo tipo sono:

- Compact Disc (CD), che hanno una capienza massima di 650 MB e che corrispondono a circa 70 minuti di musica;
- Digital Versatile/Video Disk (DVD), che possono contenere da 9 a 17 GB, con una capacità di circa 25 volte quella di un CD.

Compact Disc Sono un tipo di supporto presente sul mercato dagli anni '80. La loro diffusione è stata rapida e legata soprattutto alla memorizzazione di file audio. Per i CD le informazioni possono essere memorizzate soltanto su una faccia, l'altra faccia

è dedicata all'incisione serigrafica, alla velocità massima di 150 kB/s. È possibile suddividerlo in 3 sotto categorie che ne identificano le relative caratteristiche:

- Compact Disc-Read Only Memory (CD-ROM): è un tipo di supporto in cui è possibile scrivere soltanto una volta;
- Compact Disc-ReWriteble (CD-RW): si tratta di una evoluzione del CD Read Only Memory in quanto permette di incidere più volte la superficie del CD anche quando questo è già stato scritto, rendendo così il disco riscrivibile;
- CD-AUDIO: si tratta di un cd specifico per l'incisione di file audio per la distribuzione di musica.

Digital Versatile/Video Disk È un tipo di supporto nato dalla cooperazione di alcune fra le più grandi imprese operanti nel campo dell'elettronica di consumo, imprese come Philips, Sony, Matsushita, Hitachi, Warner, Toshiba, JVC, Thomson e Pioneer. Essi supportano fino a 240 minuti di materiale video codificato in formato MPEG-2. L'audio può avere diverse tipologie di formati come il formato non compresso PCM, il formato Dolby Digital AC3 e il formato DTS o MPEG.

Vi sono 3 tipologie di DVD:

- DVD-Video, particolarmente adatto a contenere file video;
- DVD-Audio, pensato per sostituire il CD grazie alla maggiore capienza e di conseguenza ad una maggiore qualità e fedeltà del formato originale;
- DVD-ROM, pensato per sostituire il CD-ROM.

All'interno del DVD sono stati inseriti sistemi di protezione dei dati per la copia illegale. A tal proposito vennero immessi sul mercato due tipologie di DVD: il DVD-R for authoring e il DVD-R for general use:

- Il formato DVD-R for authoring viene utilizzato per la creazione di dvd-video protetti da diritto di autore, hanno bisogno di uno speciale masterizzatore per essere scritti e possono implementare sistemi di protezione per le copie illegali;

- Il formato DVD-R for general use viene invece utilizzato come un normale dvd in cui non è possibile attuare sistemi di protezione dei dati.

La principale differenza tra i due risiede nella diversa lunghezza d'onda del laser: 635nm per quelli di primo tipo mentre 650nm per i dvd di secondo tipo. In un secondo momento sono stati immessi sul mercato altri tipi di supporti quali DVD+R e DVD+RW. Differiscono notevolmente dai supporti precedenti in termini tecnici, ma i moderni lettori e masterizzatori supportano entrambe le tipologie di formato. Come per i cd riscrivibili anche i DVD possono essere riscrivibili e sono sempre identificati dalla lettera W. Inoltre possono essere caratterizzati da un codice che identifica la regione (la parte di mondo) in cui tale cd può essere riprodotto, usando un lettore DVD da tavolo:

1. Tutto il mondo (nessuna restrizione);
2. Canada, USA e suoi territori;
3. Europa, Giappone, Hong Kong, Macao, Sudafrica, Medio Oriente, Egitto e dipendenze danesi, francesi, olandesi e britanniche;
4. Sudest asiatico;
5. Australia, Nuova Zelanda, America centrale e meridionale;
6. Russia, India, Asia nordovest e Africa del Nord;
7. Repubblica Popolare Cinese (tranne Hong Kong e Macao);
8. Riservato;
9. Speciali sedi di riunioni (aeroplani, navi, hotel, etc.).

Al contrario dei CD, i DVD implementano una seconda caratteristica. Sono infatti stati immessi sul mercato i cosiddetti DVD “*double layer*” che permettono una incisione su entrambi i lati. La capacità di supporto viene così quasi raddoppiata. Non è esattamente raddoppiata in quanto una parte dello spazio viene dedicata ad un indice di controllo della distribuzione dei dati. Altra caratteristica che differenzia i CD dai DVD è

la velocità di trasmissione dei dati che nel caso dei DVD raggiunge una velocità minima 8 volte superiore a quella di un CD, 1350 kB/s. La vita dei CD e dei DVD di tipo R (solo lettura) si colloca tra i 25 e i 250 anni, con una media di circa 100 anni. I DVD-RW hanno una vita media più breve, 25-100 anni, con un valore medio di 50 anni. Si è però scoperto che mantenendo una temperatura costante al di sotto dei 30 gradi e dell'80% di umidità, la vita media può raddoppiare.

Tipologia Supporto	Ampiezza Laser (nm)	Numero Faccie	Capacità (Gb)
CD	780	1	0,740
DVD	650	1/2	da 9,4 a 17
Blu-ray	405	1/2	da 25 a 120

Tabella 3.1: Confronto tra i diversi supporti

3.2.4 Memorie elettroniche

Sono memorie di massa molto diffuse. Vengono utilizzate nella maggior parte degli apparecchi elettronici come macchine fotografiche digitali e telefoni cellulari, o utilizzate per i nuovi dischi allo stato solido (SSD, Solid State Disk) che ormai stanno diventando lo standard di memoria di massa dei computer, soprattutto quelli portatili. La caratteristica principale di queste memorie è denominata memoria Flash. Al contrario dei suoi parenti che utilizzavano dischi magnetici per la memorizzazione delle informazioni, questo tipo di memoria utilizza una componente fondamentale chiamata transistor che viene caricata positivamente per il valore 1 e negativamente per il valore 0.

Chiave USB In inglese chiamata pen drive, ha dimensioni molto contenute e viene collegata al computer tramite la porta usb. Normalmente il suo costo aumenta con il crescere della capacità disponibile. La velocità di connessione inizialmente era molto bassa 12 Mb/s, recentemente però si è arrivati con la versione 3.0 ad una velocità di 480 Mb/s. Nei casi delle chiavette usb la velocità di scrittura

non è influenzata solo dall'interfaccia usata, ma anche dal tipo di memoria flash utilizzata, e più in particolare dalla presenza o meno di microprocessori dedicati all'interno della chiavetta. Questo tipo di struttura, penna usb + microprocessore, per il momento è dedicata ad una fascia più professionale per via degli alti costi legati alla loro progettazione.

Disco allo stato solido Abbreviato in SSD è erroneamente chiamato tale in quanto al suo interno non è presente alcun tipo di disco. La parola disco deriva dall'utilizzo che tale memoria di massa assume nello svolgere il suo compito all'interno dei computer, uguale in tutto e per tutto al suo predecessore a memoria magnetica. Non ha parti meccaniche, molto leggero e resistente agli urti. Il sistema operativo non è più collocato sui dischi magnetici ma in piccole memorie Flash allo stato solido. Questo consente di eliminare il tempo di accesso della testina al settore in cui è memorizzato il dato, e di accedere al disco in un tempo infinitesimamente minore. Anche la durata di vita del supporto è superiore rispetto ai vecchi hard disk, quasi doppia. Di contro, il rapporto MB/prezzo è decisamente molto più elevato, anche se con il passare del tempo diminuisce sempre più rapidamente. Silenziosi, rapidi, efficienti, gli hard disk SSD richiedono poca energia per funzionare. Essi non hanno bisogno di frammentazione⁶, come invece avviene per gli hard disk magnetici, perché non c'è una testina magnetica che deve saltare da settore in settore per il recupero delle informazioni. L'altra faccia della medaglia però mette in serie difficoltà gli SSD. Nei normali hard disk magnetici durante la fase di eliminazione di un dato da parte dell'utente, non vi è una reale eliminazione del dato dal settore. Il settore in cui è memorizzato il dato viene camuffato contrassegnandolo come libero, ma prima di essere realmente irrecuperabile la testina deve sovrascrivere quelle celle un numero elevato di volte in modo da renderlo così corrotto da non poter essere ricostruito nemmeno dal più sofisticato degli algoritmi. Per gli SSD questo meccanismo non vale poiché lo spazio occupato precedentemente da un file cancellato, viene messo subito a disposizione per l'ottimizzazione diventando irrecuperabile dopo pochissimo tempo in cui lo spazio diventa libero. Esistono due tipologie di memoria non volatile allo stato solido. NAND e NOR. Le due

⁶http://it.wikipedia.org/wiki/Frammentazione_informatica

tecnologie hanno obiettivi completamente differenti e di conseguenza caratteristiche tecnologiche diverse. La tecnologia più utilizzata è quella NAND in quanto la NOR, per via della sua lentezza, è un tipo di tecnologia limitata ad applicazioni che non richiedono operazioni di scrittura molto frequenti. La tecnologia NAND si suddivide in due sotto categorie: la SLC (Single-Level Cell) e la MLC (Multi-Level Cell). La tecnologia SLC memorizza un bit per ogni cella di memoria elementare mentre la MLC ne memorizza due o più. In generale la tecnologia SLC prevede 100 000 cicli di lettura/scrittura per blocco, quella MLC, invece, ne prevede 10 000 per blocco. All'interno di questi hard disk viene utilizzata la tecnologia ECC per il controllo delle informazioni. In pratica questa tecnologia aggiunge ridondanza alle informazioni memorizzate in modo da renderle più affidabili. Aumentando infatti il numero di bit, si ottiene un aumento dell'affidabilità percepita dall'applicazione e, di conseguenza, della durata dell'hard disk. Altra caratteristica importante implementata per allungare la durata di utilizzo dell'hard disk è la "distribuzione del livello di usura". Il principio di questa tecnica risiede nel non far utilizzare al disco sempre o più frequentemente solo alcune celle di memoria ma distribuirle in maniera uniforme su tutte le celle disponibili.

Scheda di memoria È una memoria di massa di piccole dimensioni in grado di memorizzare informazioni digitali anche in assenza di corrente elettrica. Questo tipo di memorie hanno un duplice utilizzo. Vengono impiegate per i salvataggi dei dati di gioco nelle console, come ad esempio la PlayStation, e per memorizzare file, soprattutto immagini e video, su periferiche portatili quali fotocamere digitali. Possono essere considerate una evoluzione dei vecchi floppy disk. La loro dimensione non ne influisce la capacità di memorizzazione, che può arrivare fino a 64 GB. Data l'enorme diffusione ne esistono di vari e differenti tipi (dimensione, forma ed interfaccia) ed alcuni di questi sono proprietari, ovvero funzionano solo sui dispositivi per i quali sono stati progettati. Anche per queste memorie sono state prodotte diversi formati:

- Compact flash, prodotto e realizzato nel 1994, si basa sullo standard PCMCIA (Personal Computer Memory Card International Association). Inizialmente

avevano una capacità di soli 200 MB ma ad oggi è stata incrementata fino a 16GB;

- Smart Media, SSFDC (solid-state floppy disk card) sviluppato da Toshiba. Molto più sottile delle compact flash, questa memoria non utilizza un controller integrato. La logica di controllo è affidata al dispositivo che funge da lettore;
- MultiMediaCard, dimensioni simili a quelle di un francobollo. La sua capacità arriva fino a 128 MB. Il suo collegamento non è formato da pin ma da componenti meccanici. La nota più importante per questo tipo di memorie è dato dalla possibilità di cifrare il contenuto;
- SD, sviluppato nel 1999, concentra le migliori caratteristiche di tutti gli altri supporti. Hanno una elevata capacità di trasferimento dato, 22,5 Mb/s, e un ridotto consumo energetico. Offrono capacità di memorizzazione fino a 64 GB e la possibilità di cifrare il contenuto.



Figura 3.3: Memorie removibili



Figura 3.4: Disco Rigido

3.2.5 Memorie olografiche

La memoria olografica è l'ultima frontiera raggiunta dalla tecnologia. Il suo utilizzo ha cambiato totalmente non solo il modo di memorizzare le informazioni, ma anche la quantità di dati che è possibile memorizzare. Viene considerata come il futuro successore del DVD e prende il nome di HVD (Holographic Versatile Disc). Ha la stessa dimensione di un normale CD o DVD ma la tecnica di memorizzazione è rivoluzionaria. Questo tipo di memoria sfrutta appieno l'intera capacità del supporto di memorizzazione grazie ad un fenomeno chiamato selettività di Bragg⁷, che permette a più ologrammi di essere memorizzati contemporaneamente nello stesso volume. Verranno create memorie a una sola scrittura e memorie che potranno essere scritte più volte. Per l'incisione del supporto saranno utilizzati due laser, il primo usato come riferimento per il secondo che si occuperà effettivamente della scrittura. Produrranno delle figure di interferenza che verranno memorizzate sul supporto. Durante la fase di lettura verrà utilizzato solo uno dei due laser. Nel momento in cui il laser attraverserà il punto in cui è stato scritto il

⁷<http://ww2.unime.it/weblab/ita/bragg/bragg2.htm>

dato, si dividerà in due, e verrà riflesso quello che risulta come fascio dati. Tale fascio dati sarà poi diretto su un dispositivo (CCD⁸) simile a quello già usato nelle macchine fotografiche, tramite il quale sarà possibile leggere le informazioni. Un aspetto molto importante è legato all'angolazione con cui i raggi incidono la superficie del supporto. Basterà infatti variare l'angolazione del raggio per poter ottenere una quantità superiore di spazio di memorizzazione. Questo tipo di supporto sarà rivoluzionario in quanto la memorizzazione dei dati non avverrà su un supporto bidimensionale ma tridimensionale. La sua velocità di trasferimento dati sarà di 160 megabit al secondo e avrà una durata stimata di 50 anni. Questi risultati sono stati ottenuti grazie all'utilizzo di una nuova tecnologia che sfrutta i fotopolimeri⁹. Per poterli utilizzare è stato necessario trovare un supporto che permettesse di conservarne la loro polimerizzazione¹⁰ poiché senza un adeguato supporto, i fotopolimeri si deformerebbero. Pavel Cheben, del National Research Council canadese, e María Calvo, dell'Università di Madrid, sono riusciti ad incidere e a conservare le informazioni per oltre sei mesi su un composto di plastica e vetro. Ciò che sarà veramente memorizzato sul supporto non sarà una sequenza di zero e uno ma una sequenza di zone di luce e ombre che saranno poi codificate dal processore come una sequenza binaria. Fino ad ora nessun altro supporto di memorizzazione ha mai raggiunto un simile risultato.

3.2.6 Memorie a singolo elettrone

Sono memorie del futuro. Non ancora realizzate ma che potrebbero in un futuro non molto lontano essere una realtà. Questo tipo di memoria sta mettendo le sue radici in un laboratorio statunitense specializzato nella microelettronica. L'idea di base di questa tecnologia è di usare uno o più slot di silicio cristallino o altro semiconduttore. Questo studio potrebbe rivoluzionare il mondo delle memorie di massa perché sarebbe possibile realizzare un chip con capacità dell'ordine di 1000 Gb con un consumo di soli 0,1 watt.

⁸http://it.wikipedia.org/wiki/Dispositivo_a_carica_accoppiata

⁹È un composto fotosensibile che diventa solido quando è esposto alla luce solare, luce ultravioletta o laser

¹⁰<http://it.wikipedia.org/wiki/Polimerizzazione>

3.2.7 Memorie organiche

Takeo Someya e Tsuyoshi Sekitani, professori e ricercatori al dipartimento Electrical Engineering and Information Systems dell'Università di Tokyo, stanno analizzando un nuovo tipo di memoria simile alla memoria flash ma con la sostanziale differenza di essere composta da materia organica. Utilizzando un foglio di resina poliestere denominato (PEN), i ricercatori hanno scoperto la possibilità di curvare tale foglio fino a 6mm senza provocare danni meccanici o elettrici. Questo substrato è composto per il momento da 26x26 celle di memoria realizzate con una struttura a doppio-transistor di tipo floating-page. Questo tipo di struttura che è già usata all'interno di altre strutture come quelle in silicio, ma insieme ad un semiconduttore organico e ad una pellicola isolante dello spessore di due nanometri, darà la possibilità di scrivere e cancellare i dati per oltre un migliaio di volte. Il limite di questa memoria è data proprio dall'attuale grandezza dello strato isolante che gli sviluppatori pensano di superare al più presto.

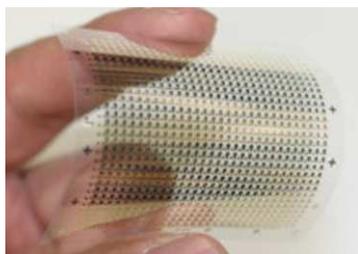


Figura 3.5: Memorie Organiche

3.2.8 HAMR

Nonostante gli hard disk allo stato solido (SSD) rappresentino il futuro, le industrie non fermano la ricerca e nuove tecnologie entreranno presto sul mercato per migliorare notevolmente le capacità dei dischi magnetici. Uno dei colossi che opera in questo settore ha annunciato una rivoluzionaria tecnologia che farà aumentare la capacità di memorizzazione di questi hard disk. Protagonista di questa scoperta è Seagate con la tecnologia HAMR (Heat-Assisted Magnetic Recording). Attualmente il limite per gli hard disk da 3,5 pollici è di 620 Gb per pollice quadrato con una capacità massima di 3 tera.. La

nuova tecnologia, promette Seagate, permetterà la densità di 1 Tb per pollice quadrato. L'HAMR punta ad aumentare la capacità di memorizzazione di ogni singolo disco all'interno dell'hard disk riducendo sulla superficie il numero di bit dei dati memorizzati e la grandezza dei cerchi concentrici. Questa tecnologia si basa su un diverso processo di memorizzazione dei bit. Attualmente i bit sono memorizzati con un procedimento definito superparamagnetismo¹¹, le particelle contenute sul disco vengono magnetizzate e polarizzate. La nuova tecnologia utilizza materiali più stabili la cui incisione richiede un laser in grado di riscaldare le particelle e di cambiarne l'orientamento. Le particelle all'interno dei dischi saranno formate da un mix di ferro e platino utilizzati al posto dell'attuale cobalto-platino.

3.2.9 Patterned Media

Un'altra tecnologia pronta a cambiare la capacità degli hard disk magnetici è stata prodotta da Hitachi, altro grosso produttore mondiale di hard disk. La tecnologia su cui punta la casa produttrice si basa sul concetto di riduzione del numero di elementi di cobalto-platino aumentandone la grandezza e isolando le singole particelle. Per quanto a prima vista possa sembrare un controsenso, i bit occuperanno meno spazio con un conseguente aumento della densità dei piatti. Al contrario della tecnologia impiegata da Seagate, c'è un grosso svantaggio. Per poter incidere la superficie bisognerebbe utilizzare un laser ad alta energia o addirittura a fasci di elettroni. Inoltre per poter sfruttare a pieno questa tecnologia bisognerebbe scendere ad una risoluzione inferiore agli 80 micron, che attualmente non è possibile ancora superare e per tal ragione è impossibile utilizzare la patterned media. Inoltre anche se si riuscisse a superare questo ostacolo bisognerebbe migliorare la testina di lettura/scrittura in modo da garantire un rapporto segnale/rumore sufficiente.

3.3 Modi organizzativi della memoria

La memoria è la parte del computer destinata alla conservazione delle informazioni, al loro recupero e alla loro manipolazione ed utilizzo. È una parte essenziale poiché ci

¹¹<http://it.wikipedia.org/wiki/Superparamagnetismo>

permette di elaborare una quantità di informazioni che il cervello umano non potrebbe mai elaborare. Nella memoria non sono solo memorizzati i dati in forma di archivio dell'utente ma anche informazioni dei programmi sfruttabili dall'utente a cominciare dal sistema operativo. Fondamentalmente tutte le memorie sono formate da un numero di celle finito che contiene una sequenza di bit, normalmente raggruppati in gruppi da 8 bit che formano un byte. Pertanto la memoria può essere vista come uno spazio fisico suddiviso in diverse locazioni nelle quali sono contenuti i byte. Per poter accedere alle informazioni queste locazioni sono identificate da un preciso indirizzo di solito espresso da un numero intero positivo.

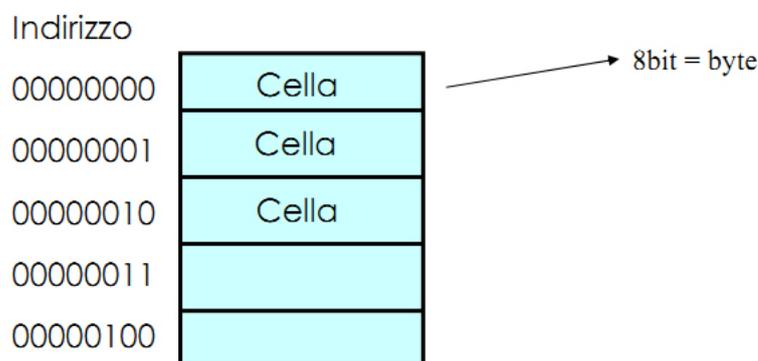


Figura 3.6: Cella di Memoria

All'interno di ogni cella la sequenza di bit rispetta un ordine ben preciso. Ogni sequenza di bit ha una estremità di ordine basso che identifica i bit meno significativi e determina se il numero è pari o dispari ed è posizionata sull'ultima casella a destra della sequenza, e una di ordine alto che identifica i bit più significativi, cioè la posizione del bit che ha il valore più grande ed è posizionata a sinistra della sequenza.

In ogni calcolatore possiamo identificare essenzialmente quattro tipi di memorie; tre fisiche: memoria centrale, memoria di massa, memoria di sola lettura, e un tipo di memoria chiamata virtuale. La memoria centrale può essere rappresentata in forma piramidale e vi si trovano in sequenza partendo dall'alto le seguenti unità di memoria: registri, cache, ram, memoria virtuale

- i registri, sono interni all'unità di elaborazione CPU e contengono operandi e risul-

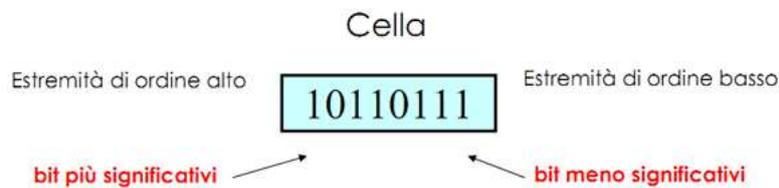


Figura 3.7: Bit

tati dell'operazione correttamente svolta dall'elaboratore. Hanno una capacità di poche centinaia di byte;

- le cache, sono piccole memorie interne collocate vicino al microprocessore; memorizzano dati temporanei usati di recente e che probabilmente verranno riutilizzati presto. Hanno una capacità inferiore ad 1 MB;
- la ram, è collocata sulla scheda madre e comunica con il processore attraverso un canale chiamato bus. Ha una velocità inferiore a quella cache e dimensioni molto elevate;
- la memoria virtuale, generalmente è uno spazio fisico del supporto di memorizzazione utilizzato come sistema di memorizzazione della memoria centrale. Ha tempi di accesso superiori a quelli della ram e può anche essere esterna al computer.

3.3.1 Accesso diretto o accesso sequenziale

Con questi due termini si indica la caratteristica di accedere ai dati dei vari supporti in un certo lasso di tempo:

- l'accesso casuale permette di accedere ad un elemento arbitrario in tempo costante e indipendente dalla dimensione del dato memorizzato. Per arbitrario si intende un elemento all'interno del supporto la cui posizione non è prevedibile. Viene utilizzato un algoritmo che determina gli indirizzi fisici che prendono come riferimento gli indicatori o chiavi secondo cui i records vengono memorizzati;
- l'accesso sequenziale indica l'accesso ad elementi memorizzati all'interno di un supporto in sequenza ordinata. La sequenza ordinata è usata tipicamente in supporti

come il nastro magnetico o quando si vuole processare una sequenza di dati che devono rispettare un dato ordine. È necessario scorrere l'intero archivio fino a quando viene trovata l'informazione che si desidera reperire.

3.3.2 Tracce e settori

I floppy disk e gli hard disk possono essere utilizzati soltanto dopo la loro formattazione. La formattazione suddivide la superficie del disco in tracce circolari e settori a spicchio tramite contrassegni elettronici in modo tale da poter archiviare e recuperare sistematicamente i file. Come possiamo vedere dalla figura 3.8 le tracce vengono identificate sul supporto da sottili fasce concentriche che nei supporti di vecchia generazione avevano una forma a cuneo che partiva dal centro del disco e ripartiva in ordine uguale la superficie. Questo tipo di suddivisione provocava uno spreco di spazio poiché le tracce più vicine ai bordi esterni erano molto più lunghe di quelle interne. Per risolvere questo problema è stata utilizzata la registrazione a zona multipla o registrazione a bit di zona. Tale tecnica permette di suddividere le tracce in un numero di settori proporzionale alla loro lunghezza. L'utilizzo di questa tecnica non ha del tutto risolto il problema in quanto non è possibile memorizzare le informazioni nelle tracce più interne, vicino al perno centrale di rotazione, così come non possono essere incise le tracce troppo vicine alla circonferenza esterna del disco. I nastri magnetici hanno una formattazione che può essere suddivisa in tre diverse modalità: longitudinale, a serpentina, a elica.

- Le tracce longitudinali sono parallele ai bordi del nastro che percorrono in tutta la sua lunghezza;
- Le tracce a serpentina sono parallele ai bordi del nastro ma al contrario di quelle longitudinali ogni traccia ha il verso opposto di quella precedente;
- Le tracce a elica sono incisioni parallele disposte diagonalmente a piccoli tratti.

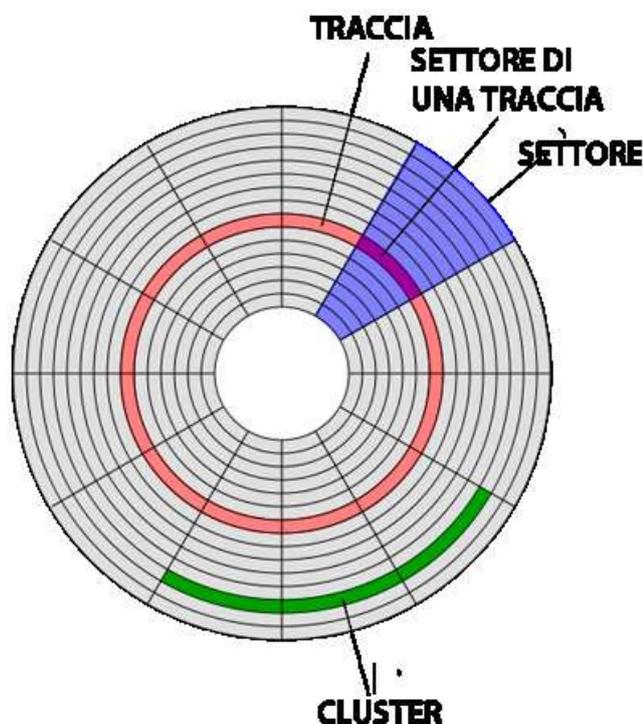


Figura 3.8: Struttura Disco

3.3.3 Velocità

La velocità di trasmissione di un supporto dipende da due caratteristiche fondamentali: la velocità di rotazione del disco, e il tempo di posizionamento. Una volta individuata la traccia i dati devono scorrere sotto la testina di lettura. Il tempo impiegato dalla testina per scorrere i dati è chiamato latenza. La latenza sommata al tempo di posizionamento della testina corrisponde al tempo di accesso, il tempo medio impiegato per iniziare la lettura dei dati. Questo tempo varia da supporto a supporto e dipende fondamentalmente dal tipo di supporto impiegato. Il più veloce è quello della memoria flash, poi vi è quello degli hard disk e infine quello dei dischetti e dei nastri. Una volta letti i dati questi devono essere trasferiti alla RAM ed in questo caso si parlerà di velocità di trasferimento dati. La velocità di trasferimento dati dipende anche in questo caso dal tipo di supporto utilizzato e dal BUS di sistema. Per velocizzare questa velocità,

nei supporti che utilizzano la testina è possibile aumentare la velocità dei dischi sotto la testina. I dischi più veloci possono arrivare ad una velocità di rotazione di oltre 15.000 rmp (rotazioni per minuto).

3.4 Porte di comunicazione

Le porte di comunicazione hanno il compito di creare un collegamento tra il supporto e il sistema hardware per poter permettere la trasmissione dei dati e la comunicazione con la memoria centrale.

3.4.1 Seriale o Parallelo

La comunicazione tra CPU e supporti di memorizzazione per lo scambio di dati può avvenire inviando singolarmente un bit alla volta in sequenza oppure contemporaneamente. Nel primo caso si parla di connessione seriale mentre nel secondo di connessione parallela. Naturalmente bisogna tenere a mente che la trasmissione tra le due unità avviene in tempi brevissimi.

Le porte seriali sono identificate dalla sigla RS-232-C (standard imposto dalla Electronics Industries Association) e sono utilizzate in particolar modo per stampanti, modem, soprattutto per periferiche definite attive come mouse, modem, scanner, penne ottiche, ed in generale per dispositivi esterni che non richiedono un flusso di dati molto veloce. I dati vengono inviati un bit alla volta e per trasmettere un byte è necessario inviare otto bit. Consente il transito dei dati in entrambe le direzioni, dal computer alla periferica e viceversa. Il connettore della porta seriale a 9 punti di contatto (PIN); uno serve per i bit di trasmissione, uno per i dati, uno per il collegamento a terra, e i rimanenti per eventuali scambi di segnali tra la periferica e il pc.

Le porte parallele sono utilizzate anche queste per alcune tipo di stampanti e per alcuni dispositivi esterni di memorizzazione. La velocità di comunicazione tra la periferica e il pc è molto più veloce rispetto alla porta seriale. In questo caso il flusso di dati avviene un byte alla volta e i bit viaggiano parallelamente su otto fili. In questo caso il transito dei bit avviene solamente in una direzione. Quando la periferica o il pc stanno comunicando viene mandato alla rispettiva parte un segnale di occupato. Il connettore

della porta parallela è formato da 25 PIN, otto dei quali per l'invio dei dati e gli altri per l'invio e lo scambio dei segnali (segnali come quelli per indicare lo stato di "occupato" o quello di "carta esaurita"). Spesso sulle porte appare la scritta Centronics (casa produttrice di questo tipo di porte che nel tempo è diventato di fatto uno standard).

Alcuni componenti esterni hanno una porta standard definita *SCSI* (Small Computer System Interface) che permette di collegare diverse periferiche in cascata fino ad un massimo di otto periferiche. Questa interfaccia viene utilizzata soprattutto per la comunicazione con unità hard disk, ma anche per scanner, lettori e masterizzatori, quindi di periferiche molto veloci, con un trasferimento dei dati che va dai 3 MB/s con la SCSI-1 ai 40 MB/s con gli ULTRA SCSI

3.4.2 USB e Firewire

Entrambe le tecnologie hanno in comune diverse caratteristiche:

- permettono di connettere in cascata diversi dispositivi, anche fino ad alcune centinaia;
- sono molto più veloci rispetto alle porte seriali e parallele;
- possono far aggiungere e rimuovere periferiche con il meccanismo definito hot plug¹² (connessione a caldo), senza cioè dover spegnere il computer;
- possono alimentare di energia elettrica le periferiche connesse.

Le porte USB (Universal Serial Bus), possono far connettere in cascata ben 127 periferiche ed in genere sono utilizzate su dispositivi che non richiedono un trasferimento di grandi quantità di dati. Attualmente la loro velocità di connessione è arrivata a 4800 Mb/s con la versione 3.0. Il collegamento Firewire, noto anche come IEEE 1394 (Institute of Electrical and Electronics Engineering, 1934), permette di collegare a cascata 64 periferiche ed è adatto alla trasmissione di una grande quantità di dati, quali ad esempio quelli di telecamere e fotocamere. È previsto un cavo per i dispositivi connessi

¹²È un'interfaccia che permette il collegamento e/o lo scollegamento di un dispositivo anche a sistema avviato

tramite usb e che abbiano un consumo minimo. Il cavo non deve superare i 5 metri altrimenti si avrebbe una diminuzione del segnale. Generalmente è usato sugli hard disk esterni per la sua praticità nel collegare e scollegare il supporto a caldo rispetto ad un collegamento di tipo ATA.

3.4.3 Wireless

Sempre più periferiche oggi godono della possibilità di sfruttare la tecnologia wireless per essere connessi al pc. Le tecnologie wireless si dividono in:

Bluetooth Viene utilizzata per il collegamento di periferiche vicine quali telefoni cellulari, o stampanti; utilizza una frequenza radio di 2,45GHz sicura e a corto raggio (da 1 a 100 metri) e con una velocità attuale di 3 Mb/s, che garantisce una comunicazione economica e sicura per la trasmissione di dispositivi tecnologicamente diversi. Può permettere la comunicazione di 7 dispositivi ma la comunicazione potrà avvenire soltanto una alla volta.

Wi-Fi Abbreviazione di Wireless-Fidelity, utilizza onde radio per connessioni a media distanza limitate da normative specifiche legate al rischio elettromagnetico. Per poter ampliare il loro raggio di azioni è necessario disporre di un access point. Si basa sullo standard IEEE 802.11. Se non vi sono ostacoli il segnale decade dopo 30 metri. Esistono diverse categorie con prestazioni diverse. Si passa dalla classe b con 11 Mb/s, alla classe g con 54 Mb/s, ad arrivare alla classe n con 450 Mb/s.

3.5 Problematiche tecnologiche

Tutto ciò che comunemente facciamo durante la giornata utilizza ormai la tecnologia digitale. Difficilmente vengono scritte lettere con carta e penna, spesso le chiamate sono effettuate tramite l'utilizzo del VOIP o ascoltiamo la musica in formato digitale tramite un MP3¹³. I momenti più importanti vengono immortalati da immagini o video in digita-

¹³È un algoritmo di compressione audio di tipo lossy, sviluppato dal gruppo MPEG, in grado di ridurre drasticamente la quantità di dati richiesti per memorizzare un suono, rimanendo comunque una riproduzione accettabilmente fedele del file originale non compresso

le e se bisogna scrivere un documento lo si fa utilizzando un word processor. Nonostante l'utilizzo di questa tecnologia sia molto semplice e permetta di risolvere innumerevoli problemi, è anche una tecnologia estremamente fragile e potrebbe facilmente scomparire se non venissero utilizzate delle strategie adeguate per prevenirne la scomparsa e per assicurare la conservazione nel lungo periodo. È facile rendersi conto che, per prevenire la scomparsa e assicurare la conservazione nel lungo periodo, sia necessario avere una apposita infrastruttura hardware e software. Ciò che è difficile tenere sotto controllo è la rapida evoluzione con cui questo sistema hardware/software si evolve. Questa infrastruttura cambia notevolmente non solo a distanza di pochi anni ma soprattutto per le varie tecnologie già presenti, spesso incompatibili. L'oggetto digitale si comporta in maniera diversa se riprodotto in un sistema per il quale non era stato progettato.

3.5.1 Deterioramento fisico e software

La rapida evoluzione dell'infrastruttura hardware/software rende obsolete tecnologie nate appena due anni prima. La continua evoluzione ci obbliga a stare a passo con i cambiamenti per evitare una drastica sostituzione dell'infrastruttura in nostro possesso. Velocità dei driver, supporti di memorizzazione, utilizzo delle periferiche di connessione e trasmissione dati come le porte USB, Tutto muta e cambia così velocemente che il cambiamento non è solo un fattore legato a differenze fisiche quali componenti più efficaci e di maggiore capacità, ma riguarda un salto di qualità delle loro funzionalità. Molti dei settori informatici oggi presenti sul mercato non sarebbero esistiti se questa evoluzione non avesse seguito tale sviluppo. Software e hardware sono strettamente connessi in quanto un nuovo software potrebbe avere difficoltà a girare su un vecchio sistema hardware anche solo per via di una periferica come quella di un nuovo hard disk o di un nuovo modulo di RAM. Tutto ciò aumenta l'obsolescenza dell'hardware e mette in evidenza il problema reale di riuscire a rendere compatibile l'utilizzo di vecchi contenuti digitali con i nuovi sistemi hardware/software.

3.5.2 Obsolescenza dei supporti

Anche i supporti di memorizzazione hanno subito e subiscono sempre più velocemente una profonda evoluzione. Basti pensare al floppy disk che ormai è solo un reperto storico o alle videocassette ben presto sostituite dai cd e anche questi ormai quasi obsoleti. La velocità con cui cambiano e si evolvono non dipende solo dalla tecnologia ma anche dalla necessità di soddisfare dei requisiti di memorizzazione dei dati da parte degli utenti. La sempre più diffusa alfabetizzazione informatica ha portato a necessità che fino a 10 anni fa praticamente erano impensabili. La velocità con cui l'informazione viaggia e si modifica, richiama l'attenzione alla sopravvivenza del contenuto digitale.

3.5.3 Obsolescenza dei formati

Anche se spesso è un fattore sottovalutato, quello dell'obsolescenza dei formati digitali non gioca un ruolo minore rispetto all'obsolescenza tecnologica, hardware o software. Spesso una azienda privata immette sul mercato un prodotto software che diviene standard. Per quanto l'impresa possa durare nel tempo, l'evoluzione, anche in questo caso, sostituirà lo standard utilizzato con un altro programma software. È difficile con il passare del tempo trovare strumenti di conversione che consentano di recuperare quei file in formati oggi accessibili. È importante riuscire a memorizzare le informazioni nel formato standard più diffuso perché è quello che nel tempo avrà maggiore possibilità di sopravvivere all'obsolescenza del formato.

3.6 Strategie per la conservazione del formato digitale

Le tecniche e le strategie che sono state utilizzate per la memorizzazione dei dati digitali sono state analizzate e studiate per far fronte al problema dell'obsolescenza delle infrastrutture hardware/software e al deterioramento fisico dei supporti. I principali studi hanno riguardato il trasferimento sui supporti analogici, conservazione tecnologica, emulazione, riversamento diretto (refreshing) e sostitutivo (migration), archeologia digitale.

3.6.1 Trasferimento sui supporti analogici

La prima strategia utilizzata per la memorizzazione dei dati digitali va, in una prima analisi, in controtendenza a quella che è la caratteristica della digitalizzazione. Si tratta infatti di stampare su carta informazioni digitali che generalmente sono da anni utilizzate come standard. Questa strategia era utilizzata per tutti quei formati analogici fragili, per i quali si procedeva oltre alla fase di riparazione anche alla creazione di una copia, ad alta fedeltà, dello stesso materiale. Questo ha permesso di soddisfare anche le richieste di utilizzo e di accesso a tali documenti preservando quelli originali. Vi sono però dei problemi a monte dell'utilizzo di questa strategia. Un primo fattore riguarda il fatto che molti documenti, al giorno d'oggi, nascono direttamente in formato digitale. Il secondo fattore è legato al fattore di duplicazione dei documenti digitali, che in molti casi, non è possibile trasferire su un documento cartaceo, come ad esempio documenti audio e video o email. Come trasferire in formato cartaceo un allegato email o come preservare le caratteristiche di un collegamento ipertestuale? Il trasferimento dal digitale al cartaceo è destinato dunque a non essere una soluzione affidabile poiché oltre ad un problema di preservazione dei dati digitali c'è un problema più realistico che dipende dalla quantità di dati digitali attualmente in circolazione.

3.6.2 Conservazione tecnologica

Altra strategia che sembra adattarsi meglio alla risoluzione del problema riguarda la conservazione tecnologica, più precisamente quella software (sistema operativo, programmi, driver) e hardware. Questa soluzione porterebbe alla creazione di veri e propri musei del computer nei quali conservare in buono stato tutti i computer con i relativi programmi e le periferiche necessarie che si intendono conservare. Una tale strategia risulta sicuramente più appropriata rispetto al trasferimento in formato cartaceo delle informazioni ma, anche in questo caso il problema riguarda il mantenimento in funzione di tutte queste tecnologie all'infinito. Questo porta ad investimenti finanziari troppo alti da sostenere nel tempo. Quindi è un tipo di strategia ritenuta valida solo per tecnologie considerate veramente importanti e rare.

3.6.3 Emulazione

Questa strategia ideata da Jeff Rothenberg, si basa sul concetto che l'informazione digitale possa essere garantita emulando il documento digitale su sistemi tecnologicamente più evoluti. Si tratta di allegare oltre al documento digitale, anche il sistema operativo, il software utilizzato per la sua creazione, i metadati, ed un emulatore della piattaforma hardware nativa. Anche in questo caso vi è l'altra faccia della medaglia che individua come questa procedura non sia semplice da attuare. In particolare David Bearman sostiene che hardware e software siano di norma proprietari e quindi difficilmente incapsulabili a causa dei diritti detenuti dai relativi creatori e che vi siano difficoltà di lettura dei metadati su architetture future e non conosciute, in quanto hanno bisogno di una architettura ben definita per una loro corretta implementazione.

3.6.4 Riversamento diretto (refreshing) e sostitutivo (migration)

È la strategia digitale che volge lo sguardo al singolo bit. Infatti con il passare del tempo e a seconda del supporto di memorizzazione utilizzato, la sequenza di bit potrebbe cambiare, perdendo progressivamente alcuni bit. L'idea è quella di copiare (refreshing) i dati digitali, in periodi brevi di tempo, attraverso il refreshing su supporti (hardware e software) di volta in volta più recenti. Il refreshing in particolare consiste nel copiare i dati digitali da un supporto di memorizzazione obsoleto ad un nuovo supporto senza modificare la sequenza di bit originale. Per poter far ciò il nuovo supporto deve essere dello stesso tipo o simile. Questa strategia è relativamente semplice e poco rischiosa. Oltre questa strategia bisogna anche utilizzare quella denominata riversamento sostitutivo (o migration), che consiste nell'operazione opposta al riversamento diretto. In questo caso la memorizzazione dei dati digitali da un vecchio supporto ad uno nuovo prevede una ricodifica periodica dei bit, mantenendo gli aspetti principali e modificando quelli secondari per adattare meglio il dato digitale al nuovo tipo di supporto hardware e software.

3.6.5 Archeologia digitale

Bisogna fare attenzione a non confondere questa strategia con la strategia di conservazione tecnologica. Ciò che la differenzia è che l'archeologia digitale si presta a recuperare un contenuto digitale che non è stato sottoposto ad un processo di conservazione digitale e che contiene informazioni digitali di vitale importanza, oppure che nonostante sia stato sottoposto ad una conservazione, non sia improvvisamente più accessibile, o ci sia stato un deterioramento del supporto di memorizzazione. In genere è una strategia che viene adottata in casi estremi e solo quando non è possibile recuperare le informazioni in altro modo. Questo perché il recupero dei dati attraverso questa tecnica presuppone l'utilizzo specifico di tecniche di volta in volta diverse per i vari tipi di contenuti digitali, ricorrendo a società specializzate.

Capitolo 4

Backup e sicurezza informatica

L'attività di conservazione dei dati o per meglio dire il backup, è una parte fondamentale dell'informatica. Basti pensare alla quantità di dati oggi presente in formato digitale, ai dati sensibili di banche ed enti governativi, o anche semplicemente ai dati digitali conservati sul proprio hard disk personale di casa. Che si tratti di privato o aziende il problema del backup e del recupero di informazioni non esula nessuna delle parti ad attuare strategie e tecniche per poter recuperare, in casi estremi tali dati e per proteggerli da attacchi volti a prendere possesso delle informazioni. Non esiste un sistema di backup e archiviazione definitivo poiché bisogna sempre verificare le proprie esigenze, il budget a disposizione, la quantità di dati da conservare e mettere al sicuro. Inoltre non vi sono ancora oggi politiche in vigore riguardanti la sicurezza e la copia dei dati. Solo negli ultimi anni politiche di backup sono state implementate sotto precisi ordini di legge per tutto ciò che riguarda i dati i sensibili. In particolare citando la normativa italiana sul backup è stato scritto nell'articolo 15, comma 1, legge 675/96:

I dati personali oggetto di trattamento devono essere custoditi e controllati, anche in relazione alle conoscenze acquisite in base al progresso tecnico, alla natura dei dati e alle specifiche caratteristiche del trattamento, in modo da ridurre al minimo, mediante l'adozione di idonee e preventive misure di sicurezza, i rischi di distruzione o perdita, anche accidentale, dei dati stessi, di accesso non autorizzato o di trattamento non consentito o non conforme alle finalità di raccolta.

Spesso i dati sono salvati sulla stessa macchina di lavoro senza adottare particolari strategie per tutelarsi da eventuali disastri.

I motivi per cui attuare una politica di backup sono diversi:

- Usura e danneggiamento dei supporti di memorizzazione;
- Intenzionali, quando file e programmi vengono deliberatamente eliminati da un utente;
- Non intenzionali, quando i file sono erroneamente cancellati o sono provocati da un errore dell'amministratore del sistema;
- Guasti, derivanti dall'interruzione della corrente elettrica, guasto dell'hardware o crash del programma che si sta utilizzando;
- Disastro, provocato da alluvioni, incendi o terremoti;
- Crimine, dovuto al furto del supporto di memorizzazione o attraverso sistemi remoti via web con l'utilizzo di virus e trojan.

Il *disaster recovery* è diventato un requisito fondamentale ed ha influenzato enormemente l'architettura dei sotto sistemi di memoria. Più precisamente per disaster recovery si intende l'insieme di misure tecnologiche e di processi organizzativi atti a ripristinare sistemi, dati, infrastrutture necessarie per far fronte a gravi emergenze (incendi, allagamenti, furti, attacchi di virus, guasti hardware, errori umani). Per riuscire a prevenire un *disaster recovery* sono state implementate due proprietà: MTBF e useful service life. MTBF (Mean Time Between Failures) è un parametro di affidabilità applicato ai dispositivi meccanici, elettrici ed elettronici. Più precisamente è il valore atteso tra un guasto e quello successivo. Intorno al 1990 il valore atteso degli hard disk era di 200 000 h, ma nel corso degli anni grazie all'avanzamento della tecnologia si è passati a circa 1 200 000 h. Tuttavia il MTBF deve essere utilizzato insieme all'useful service life (vita media utile) per una più precisa diagnostica e prevenzione. Possono comparire nel tempo guasti o consumi delle parti meccaniche ed elettroniche. La combinazione di questi due parametri fa giungere alla conclusione che la vita media degli hard disk arriva ad appena cinque anni. Ecco perché è necessario proteggere i dati con soluzioni che consentano di sfruttare al meglio tutte le tecnologie disponibili.

4.1 Backup

Se si dispone di un computer, soprattutto di quelli più recenti, l'attività di backup è impostata in maniera automatica attraverso un software preinstallato e con una certa periodicità stabilita. In genere questa attività nelle aziende è regolata da una procedura specifica e soggetta ad una verifica periodica. I supporti su cui viene effettuato il backup devono essere approvati nella procedura ed è necessario sostituire i supporti se vecchi o danneggiati e verificare che le procedure siano efficaci. Recuperare file danneggiati è una operazione sempre molto complicata e che dipende dal tipo di danno subito. In genere un buon backup riuscirà a recuperare la totalità dei dati persi ma questo dipende molto anche dal tipo di backup e dalla procedura utilizzata. Per quanto dovrebbe essere la prima regola da seguire, spesso e volentieri il backup è una delle questioni che più viene lasciata in secondo piano e che nel momento in cui si verifica il danno si è costretti a rivolgersi ad imprese specializzate molto costose. Ecco perché è importante fare periodicamente delle copie di backup dei dati importanti che devono essere ben custodite. A seconda di quali dati e di quanti dati bisogna giornalmente salvare esistono varie tipologie di backup che è possibile utilizzare:

- backup completo è la copia di tutti i dati che si vogliono preservare, indipendentemente che vengano modificati o meno. Ad ogni operazione di backup completo tutti i dati verranno copiati;
- backup differenziale si basa su un algoritmo che confronta i file e le directory del backup precedente con quello che si dovrà eseguire copiando solo i file e le directory che hanno subito delle modifiche. Al suo primo avvio sarà effettuato un backup completo, successivamente solo i file modificati;
- backup incrementale è la copia di tutti i file che sono stati modificati dall'ultimo backup completo o differenziale. Anche in questo caso sarà effettuato come prima volta un backup completo e solo nei backup successivi i file solo modificati, ma a differenza del backup differenziale verranno mantenute anche le versioni vecchie dei file già copiati.

4.2 Strategia di backup

L'obiettivo del backup è quello di garantire un recupero dei file che sia il più veloce possibile e soprattutto con la minor perdita possibile dei dati. La pianificazione del backup, soprattutto per le grosse imprese deve avvenire pianificando una precisa sequenza di procedure. Sarebbe buona norma utilizzare diversi supporti per le varie fasi di backup dei dati e utilizzare un tipo di archiviazione off-site. Questa archiviazione si basa sulla strategia di collocare la copia di backup in un luogo distante dall'originale, e questo per ovvi motivi: per fare un esempio basta che ci sia un incendio all'interno dell'edificio e tutti gli sforzi fatti per salvare i dati non sarebbero serviti a niente. I supporti su cui memorizzare l'informazione hanno caratteristiche che li differenziano per qualità, tipo di memorizzazione e quantità che possono memorizzare. È per questo che si consiglia alle imprese di utilizzare diversi supporti per i diversi backup che devono essere effettuati. È consigliabile usare un supporto per il backup giornaliero, uno per quello settimanale/mensile, uno per un backup di tipo completo e uno per un backup di tipo differenziale o incrementale. La ragione di questa scelta è legata al fatto che ogni supporto ha la sua caratteristica di durata e di performance e soprattutto legata anche al processo di ripristino su file system in quanto il ripristino dal backup differenziale e da quello incrementale è molto differente. Nel caso di backup differenziali bisognerà:

- ripristinare i dati contenuti nel backup più completo e più recente;
- ripristinare i dati dell'ultimo backup differenziale effettuato.

Nel caso invece di backup incrementali bisognerà:

- ripristinare i dati contenuti nel backup più completo e più recente;
- ripristinare tutti i dati dei backup incrementali effettuati dopo l'ultimo completo backup, a partire dal meno recente sino all'ultimo effettuato.

Il backup di tipo differenziale effettua una copia maggiore del numero di dati che comporta un rallentamento nella fase di backup, ma dall'altro canto ha una ridondanza maggiore dei dati e una maggiore semplicità e velocità di ripristino. Il backup di tipo

incrementale invece ha tempi e complessità di ripristino maggiori ma ha il vantaggio di fornire un backup più veloce con una occupazione di memoria minore. Non vi è nessuna differenza che si basa sulla maggiore protezione dei dati. Le differenze sono basate esclusivamente sulla quantità di spazio utilizzata e sulla velocità di ripristino oltre che differenza nei costi.

4.3 NAS

Fondamentale per il disaster recovery è stato l'avvento del Network-Attached Storage (NAS) che offriva capacità di Network File System. Questo supporto è stato sviluppato nel 1984 dalla Sun Microsystems con lo scopo di utilizzare la rete per accedere ai dischi remoti come fossero dischi locali. Permette di gestire numerosi protocolli come il CIFS (Common Internet File System), L'HTTP (HyperText Transfer Protocol), FTP (File Transfer Protocol) ed è basato sul sistema operativo Linux. La particolarità di questo supporto è di permettere l'accesso ai file contemporaneamente su più piattaforme, come ad esempio Linux, Windows, Mac OSX. Implementa al suo interno il sistema RAID (Redundant Array of Independent Disks), i quali garantiscono una migliore gestione della sicurezza dei dati. Inoltre consente una eventuale rimozione a caldo (hot swap) dei dischi.

4.4 SAN

Storage Area Network (SAN) è stata creata per rendere disponibile le risorse di storage per qualsiasi architettura hardware e software. È costituita da dispositivi di memorizzazione di massa, spesso anche differenti, e da una parte di rete ad alta velocità, tipicamente (Gb/s). Questa tecnologia ha introdotto nuove possibilità di semplificazione delle connessioni, nella crescita, nei costi e nella gestione delle memorie di massa. Con essa anche la fibra ottica è cresciuta diventando la tecnologia predominante per le reti. Grazie proprio alla crescita della rete e alla diffusione della rete Ethernet e protocollo TCP/IP la SAN ha avuto un grosso sviluppo in fatto di standardizzazione dei sottosistemi di memoria. SAN lavora per garantire a qualsiasi server della rete, a cui appartiene, la disponibilità di memorizzazione dei dati garantendo una efficiente organizzazione dello

spazio e l'automatizzazione di numerose funzioni di gestione. In particolare i metadati sono affidati ad un server specifico che ne semplifica la gestione e ne migliora la disponibilità. Il vantaggio in assoluto più significativo è la possibilità di estendere geograficamente l'utilizzo di questa architettura e di creare le cosiddette SAN "metropolitane" utilizzando tecnologie trasmissive come la WAN per creare una vera e propria *global storage network*.

4.5 Innovazioni, hard disk, server virtuali, Cloud storage

Fino ad ora ciò che ha influenzato maggiormente il recupero dei file dai supporti di memorizzazione è stato l'utilizzo di memorie di massa lente, come i nastri magnetici. I sistemi di backup basati su tale tecnologia richiedono ore per un ripristino totale dei dati, oltre al fatto che tutti gli utenti devono cessare la loro attività prima di poter eseguire un backup. Questo implica che tutto il procedimento deve essere effettuato durante la fase notturna e nei week end lasciando il processo nelle mani dell'elaboratore sperando che non vi siano complicazioni durante il processo. Inoltre nonostante il nastro sia un componente poco costosa, va sostituito spesso e di conseguenza il costo dello storage aumenta. L'attuale tecnologia ha portato grosse innovazioni anche in questo campo consentendo di realizzare apposite strutture hardware e software in grado di superare questo limite. In particolar modo la nuova tecnologia fa uso del disco. La rivoluzione portata dall'utilizzo del disco permette di creare copie di backup non una volta al giorno o una volta alla settimana, ma quasi in tempo reale, grazie anche al fatto che i nuovi software consentono di salvare solo il file effettivamente modificato. I dati compressi e duplicati vengono salvati come una immagine che include anche il sistema. È chiaro che questa strategia consente oltre ad un risparmio di circa il 60% dello spazio a disposizione anche un ripristino dei dati è molto più veloce. Nel caso ci sia la necessità di un ripristino totale del server, il sistema può tornare all'immagine precedente catturata nell'ultimo istante utile. L'utilizzo di server virtuali che replicano fedelmente il sistema, consente di ripristinare il lavoro quasi all'istante. Ma ciò che sembra realmente essere il futuro dello storage, e non solo, è il sistema Cloud storage. Questo servizio offre un sistema di archiviazione simile al classico storage ma con la differenza che i file non

saranno più posizionati all'interno di un disco locale bensì, verranno salvati su un disco presente in un server (o più server) posizionato in qualche parte del mondo. Questo tipo di storage funziona anche offline grazie ad un sistema di sincronizzazione. La nuova strategia di backup presenta notevoli vantaggi ma anche notevoli svantaggi come ad esempio la possibilità di avere sempre a disposizione i dati in qualunque parte del mondo ci si trovi, l'importante è disporre di una connessione ad internet, ma uno spazio al tempo stesso limitato ed un costo aggiuntivo di GB molto più alto rispetto ad un dispositivo fisico, una condivisione e memorizzazione facile ma una velocità di trasferimento molto più bassa di una connessione USB 2.0. Il cloud storage promette però di essere una più promettente e flessibile scelta rispetto alla strategia tradizionale, soprattutto per i casi d'uso di storage dedicati a video online, audio e libri elettronici. Questo sistema è in continua evoluzione e risulta difficile poter prevedere quale sarà il suo futuro nei prossimi anni. Per rispondere a tutte le esigenze il cloud storage si è suddiviso di tre sottocategorie:

- l'implementazione pubblica è un ambiente molto sicuro e multi-tenant¹ che esternamente è a disposizione di tutti gli utenti
- una implementazione privata è un ambiente sicuro a singola entità, sia all'interno che all'esterno delle reti private
- una implementazione ibrida è una combinazione di cloud pubblici e privati.

Il cloud storage in altre parole è uno storage virtualizzato on demand, chiamato più formalmente DaaS² (Data Storage as Service). Ad oggi il cloud storage viene implementato unicamente via web, tramite una interfaccia chiamata REST³ e tramite API⁴ fornite direttamente dai provider del servizio. Inoltre il Cloud Data Management Interface (CDMI) dalla fine del 2011 è stato sottoposto all'ISO per la certificazione come

¹Utilizzato per descrivere le architetture delle applicazioni progettate per supportare diversi utenti

²È un servizio web appartenente alla tecnologia del cloud computing, che mette a disposizione dell'utente i dati rendendoli disponibili in vari formati ed ad applicazioni diverse come se fossero presenti sul disco locale

³È un tipo di architettura software per i sistemi di ipertesto distribuiti come il World Wide Web

⁴http://it.wikipedia.org/wiki/Application_Programming_Interface

standard internazionale. Questa interfaccia è stata progettata per avere una piena compatibilità con tutte le offerte di storage pubblico quali portabilità, conformità, sicurezza dei dati. Inoltre garantisce la connessione di cloud appartenenti a provider diversi garantendo la massima compatibilità. La semantica di questa interfaccia è semplice. Formata da bucket e data object codificati con metadati che ne descrivono i requisiti dell'oggetto o del contenitore. Il protocollo di accesso è il RESTful HTTP⁵. L'innovazione portata da questa nuova tecnologia e strategia sta riscuotendo un enorme successo e nonostante i limiti legati alla sicurezza, ai costi e alla velocità di trasferimento, sembra rappresentare un futuro ormai certo.



Figura 4.1: Schema di un sistema cloud

⁵È un protocollo riguardante l'esposizione di sorgenti dati via HTTP

Conclusioni

Il percorso seguito in questa tesi ha messo in evidenza i problemi in apparenza trascurabili della memorizzazione dei dati digitali. Il fatto stesso che i supporti di memorizzazione non sono perenni, indistruttibili, ma come tutti gli oggetti hanno una loro deperibilità, hardware e software, ci porta a focalizzare l'attenzione sulla prevenzione e sugli accorgimenti che gli utenti devono adottare in materia di dati digitali. Nei primi due capitoli è stata posta l'attenzione su come preservare i dati digitali onde evitare di non poterli più utilizzare nel corso degli anni. In particolar modo ci siamo concentrati sul tipo di formato utilizzato per la conservazione e sulla loro gestione ponendo l'attenzione sulla qualità e non sulla quantità dei dati da memorizzare. Come è stato più volte marcato nella tesi, la conservazione dei dati parte da una attenzione nel produrre i dati stessi nella forma più corretta possibile, attraverso l'utilizzo di procedure ben definite e di formati standard che possano essere il più possibili semplici e compatibili con la maggior parte dei sistemi e dei supporti di memorizzazione. Abbiamo visto l'utilizzo di standard sia in materia di creazione del dato, sia di attenzione alle linee guida della sicurezza informatica e solo alla fine standard per i supporti di memorizzazione. Questa fase garantirebbe la migliore gestione del materiale e di conseguenza garantirebbe una conservazione a lungo termine degli stessi dati indipendentemente dal componente hardware in cui sono memorizzati. L'attenzione poi è stata posta sui diversi dispositivi e sulla loro evoluzione. In particolare abbiamo potuto notare che nonostante siano passati anni e ci siano state parecchie evoluzioni, l'uso di dispositivi come il nastro magnetico, è tutt'oggi un elemento ben presente nella maggiorparte dei sistemi di backup. Gli stessi dischi magnetici o hard disk, nonostante sembrava avessero già concluso il loro percorso poichè mostravano limiti quali la velocità di trasmissione e la grandezza totale

di memorizzazione, hanno subito una grossa ribaltata con la scoperta di tecniche come l'HAMR che permetterebbero di alzare i limiti attuali e lanciare in avanti di molti altri anni ancora il loro utilizzo. O ancora i dischi allo stato solido o memorie flash, che stanno permettendo di avere prestazioni molto alte ma al tempo stesso pongono problemi circa il costo, dieci volte più alto di un normale hard disk, ed una minore durata dell'unità causata dal limite di riscritture delle memorie flash. Per non parlare inoltre del problema del recupero delle informazioni che nei nuovi hard disk allo stato solido sarebbe quasi impossibile da attuare con una ripercussione su tutto quello che coinvolge il settore del disaster recovery. Le momerie del futuro invece sono ancora in fase di elaborazione e per quanto promettenti non ci consentono ancora di poter azzardare una vera e propria previsione su loro futuro visto che tra l'altro abbiamo ancora tanto da imparare sugli attuali dispositivi dei quali, non siamo ancora in grado di dare con certezza parametri riguardo la loro vita. Le nuove tecniche di storage infine, stanno subendo una radicale innovazione grazie alle nuove tecnologie, in particolare con l'utilizzo degli hard disk al posto dei nastri magnetici e del WEB, in particolare del cosiddetto CLOUD Storage che offre la possibilità di sfruttare le risorse della rete e dei calcolatori connessi per poter avere uno storage sempre aggiornato e pronto all'utilizzo con un ripristino quasi immediato e la possibilità di avere sempre accesso a tutte le informazioni da qualunque parte del mondo ci troviamo. La conclusione è che non è possibile dare per certa una soluzione che utilizza soltanto una tecnologia in quanto ogni dispositivo avrà le sue caratteristiche peculiari che potranno essere fondamentali per non incorrere in brutte sorprese. Le nuove tecnologie tendono a sostituire con sempre maggiore rapidità tutti gli apparati tecnologici per la conservazione e l'elaborazione dei dati e questo è fonte da un lato di una maggiore prestazione, dall'altra è causa di una obsolescenza fin troppo rapida che potrebbe causare l'illeggibilità dei dati stessi. Certo è che il futuro dei dati digitali dipende da numerosi fattori e in primis dagli stessi utenti i quali volenti o nolenti sono i veri responsabili della corretta gestione dei file digitali.

Riferimenti bibliografici

Bibliografia cartacea

- Alberto, Salarelli e Anna Maria Tammaro: *La biblioteca digitale*, Milano: Editrice Bibliografica, 2006.
- Allegrezza, Stefano: *Informatica di base. Conoscere e comprendere le risorse digitali nella società dell'informazione*, Macerata: Edizioni Simple, 2009.
- Betti, Renato: *Analogicodigitale*, Torino: Einaudi, 1977.
- Ciotti, Fabio e Gino Roncaglia: *Il mondo digitale. Introduzione ai nuovi media*, Roma-Bari: Gius. Laterza & Figli, 2000.
- Console, Luca e Marina Ribaudò: *Introduzione all'informatica*, Torino: Utet, 2004.
- Curtin, Dennis P.: *Informatica di base*, Milano: McGraw-Hill, 2008.
- Galluzzi, Paolo e Pietro A. Valentino: *I formati della memoria. Beni culturali e nuove tecnologie alle soglie del terzo millennio*, Milano: Giunti, 1997.
- Pace, Andrew K.: *The ultimate digital library: Where the new information players meet*, Chicago: ALA Editions, 2003.

Siti Web consultati

- AA. VV.: *Osservatorio sul Documento Digitale*, 2006, URL: <http://www.infocamere.it/doc/osservatoriowhite.pdf>.
- ANORC: *Nuova versione degli standard ETSI*, 2012, URL: http://www.anorc.it/notizia/288_Nuova_versione_degli_standard_ETSI.html.
- Auletta, Vincezo: *Le memorie secondarie*, 2004, URL: <http://www.dia.unisa.it/professori/auletta/DIDATTICA/INFORMATICA/slide/memoriaMassa.pdf>.
- Aversano, Lerina: *La memoria di massa. La gestione dell'IO*, 2007, URL: www.rcost.unisannio.it/aversano/Lucidi/lezione8.pdf.
- Bazzo, Carlo: *Campionamento e quantizzazione di immagini*, 2011, URL: <http://www.hdemo.com/digital-imaging-theory/campionamento-e-quantizzazione-di-immagini/>.

- Bentivoglio, Gianfranco: *Informatica*, 2005, URL: ebookbrowse.com/2478-teoria-cap-3-2-hw-memorie-di-massa-rev-2.
- Bergamin, Giovanni: *La conservazione delle risorse digitali*, 2005, URL: <http://www.rinascimento-digitale.it/formazione/corso-conservazione-giugno2008/bergamin-cosasipuofare.pdf>.
- Caratti, Marco: *Le memorie a semiconduttore*, 2006, URL: http://www.ilb2b.it/sites/default/files/_importazione/automazione_oggi/pdf_riv/200608/20060829005_1.pdf.
- Cilli, Antonio: *Il documento informatico e la firma digitale*, 1998, URL: <http://giuriweb.unich.it/docinfo1.html>.
- Claudio56: *Hard disk: le nuove tecnologie intralciano le indagini della polizia*, 2011, URL: <http://claudio56.wordpress.com/2011/03/08/hard-disk-le-nuove-tecnologie-intralciano-le-indagini-della-polizia/>.
- De Robbio, Antonella: *Metadati per la comunicazione scientifica*, 2001, URL: <http://www.iccu.sbn.it/upload/documenti/Derobbio.pdf>.
- Di Domizio, Alessio: *La dura vita dei dati digitali*, 2008, URL: <http://www.appuntidigitali.it/1601/la-dura-vita-dei-dati-digitali/>.
- Di Iorio, Angela: *Depositi digitali e problemi di conservazione*, 2009, URL: <http://illinguaggiodellebibliotechedigitali.pbworks.com/f/Appunti.htm>.
- Doria, Gianni Penzo: *La conservazione del documento digitale*, 2006, URL: <http://www.interlex.it/docdigit/gpenzo2.htm>.
- Ebranati, Valentina: *Il protocollo Z39.50*, 2004, URL: <http://www.nmis.isti.cnr.it/savino/Esami/Primo%20appello/Il%20protocollo%20Z39.50%20Valentina%20Ebranati.pdf>.
- Foscarini, Fiorella: *I metadati per la gestione e conservazione dei documenti elettronici*, 2007, URL: <http://www.rinascimento-digitale.it/formazione/corso-conservazione-giugno2008/foscarini-2-metadati.pdf>.
- Franchi, Daniele: *Backup e Sicurezza*, 2008, URL: http://131.114.106.134:8080/TESTNET/presentazioni_ppt/fah11/3/Backup%20e%20sicurezza.pdf.
- Granduglia, Andrea: *Scrivere in codice. Il testo al tempo del Web*, 2000, URL: http://www.noemalab.org/sections/specials/tetcm/scrivere_in_codice/main.html.

- Guadagni, Franco: *L'obsolescenza dei supporti tecnologici: "Quanto durano i ricordi digitali?"*, 2006, URL: http://www.telecomitalia.com/TIPortale/docs/innovazione/012006/Ricordi_digitali_1_2006.pdf.
- Guercio, Maria: *La conservazione a lungo termine dei documenti elettronici: normativa italiana e progetti internazionali*, 2001, URL: <http://archive.forumpa.it/archivio/0/200/260/263/GuercioConservazioneDeiDoc.pdf>.
- Hofmann, Ernesto: *L'evoluzione delle memorie periferiche*, 2005, URL: <http://ebookbrowse.com/hofmann-p-17-31-pdf-d89425497>.
- Kosmous: *Backup e restore*, 2003, URL: <http://www.kosmous.com/manuali/Backup-e-restore-I?keyc=34327727FCFDC61D63BE733EADBBBDB5&cm=0>.
- Kryder, Mark H.: *Heat Assisted Magnetic Recording*, 2008, URL: <http://nanomag.ucsd.edu/wp-content/uploads/2011/12/Kryder-review.pdf>.
- Lattanzi, Emanuele: *Memorizzazione dei dati*, 2005, URL: www.sti.uniurb.it/lattanzi/SCMOT/lezioni/Lez4.pdf.
- Magliano, Cristina: *Lo standard nazionale dei metadati gestionali amministrativi*, 2006, URL: http://digitalia.sbn.it/upload/documenti/digitalia20050_MAGLIANO.pdf.
- Messina, Maurizio: *La conservazione dei contenuti digitali*, 2005, URL: http://www.bncf.firenze.sbn.it/bolzano/M_Messina.pdf.
- Michetti, Giovanni: *Il modello OAIS*, 2010, URL: <http://www.rinascimento-digitale.it/formazione/corso-conservazione-giugno2008/michetti-modello-oais.pdf>.
- Microsoft: *Dove é consigliabile salvare il backup?*, URL: <http://windows.microsoft.com/it-IT/windows7/Where-should-I-save-my-backup>.
- Redazione HostingTalk: *Cloud Storage, una grande opportunità per lo storage del futuro*, 2011, URL: <http://www.hostingtalk.it/articoli/cloud-computing/9745/cloud-storage-una-grande-opportunita-per-lo-storage-del-futuro?page=5>.
- Santoro, Michele: *Dall'analogico al digitale: la conservazione dei supporti non cartacei*, 2001, URL: www.bibliotecheoggi.it/2001/20010208801.pdf.

- Savino, Nicola: *Lo standard europeo per i software e i sistemi di conservazione sostitutiva*, 2011, URL: <http://www.nicolasavino.com/2011/06/08/lo-standard-europeo-per-i-software-e-i-sistemi-di-conservazione-sostitutiva/>.
- Sbiaa, Rachid and Piramanayagam, Seidikkurippu N.: *Patterned Media Towards Nanobit Magnetic Recording: Fabrication and Challenges*, 2006, URL: <http://ebookbrowse.com/piramanayagam-pdf-d83177763>.
- Trucco, Paola: *Tutta la memoria in un solo elettrone*, 2001, URL: www.cnr.it/documenti/Istituzionali/Consuntivo/pdf/122125fo.pdf.
- Valacchi, Federico: *Il documento informatico*, 2010, URL: <http://docenti.unimc.it/docenti/federico-valacchi/>.
- Vitali, Fabio: *Il markup di documenti*, 2003, URL: <http://www.cs.unibo.it/~fabio/corsi/tw02/slides/08-Markup/>.

Ringraziamenti

Questa è la conclusione di un percorso che in tutti questi anni mi ha portato a raggiungere non solo la conclusione dei miei studi universitari, ma soprattutto mi ha permesso di crescere e di raggiungere molti altri importanti obiettivi della mia vita. In particolare i miei ringraziamenti vanno alla mia famiglia ma soprattutto a mia mamma che non mi ha mai fatto mancare nulla grazie ai suoi sacrifici e che ha sempre creduto in me lasciandomi la libertà di vivere la mia vita e il mio futuro. A Katia, prima una grandissima amica e poi compagna di vita e di viaggio. Una persona unica di quelle che è raro trovare nella vita e che, fin dal primo incontro, è stata sempre accanto a me sostenendomi in tutte le mie scelte. Ai miei migliori amici con i quali ho imparato a superare tante difficoltà e che mi hanno aiutato a capire come è sempre possibile raggiungere tutti gli obiettivi quando ci si mette la passione e il cuore in quello che si fa. A Lorenzo&Stella con cui ho condiviso un ultimo percorso importante che siamo riusciti a superare insieme.

A mio padre...