



Corso di

Progettazione e gestione di

risorse digitali per i beni

culturali

prof. Pierluigi Feliciati

a.a.2012/13

Modulo 1

Digitale, definizioni, formati, supporti delle risorse digitali

Le aggregazioni di risorse: database, web, collezioni, sistemi informativi

Per iniziare: cos'è l'informazione?

Un possibile percorso: la transizione del business

L'evoluzione economico-industriale della società occidentale può essere interpretata secondo tre fasi:

1 - Dapprima l'attenzione è stata rivolta alla produzione di beni materiali, nella società pre-industriale: **gestione della materia**

... producendo beni materiali, ci si è anche resi conto dell'importanza di produrre in modo efficiente, nella

2 - società industriale: **gestione dell'energia**

... impiegando energia, ci si è resi conto dell'importanza di controllare adeguatamente la produzione, nella

3 - società post-industriale: **gestione dell'informazione**

cos'è l'informazione?

la transizione delle reti

Le reti da sempre costituiscono per la società uno strumento indispensabile di trasporto, scambio e comunicazione:

- dapprima reti per consentire il **trasferimento di persone e beni materiali**: reti stradali, reti fluviali, acquedotti, reti fognarie, ...
- quindi anche reti infrastrutturali per consentire il **trasferimento di energia**: reti elettriche, gasdotti, reti petrolifere, ...
- quindi anche reti per consentire il **trasferimento di informazione**: reti telegrafiche, telefoniche, televisive, informatiche, ...

informazione e mondi

informazione fisica, conoscenza e rappresentazione



con l'introduzione dei **computer**, si è arrivati a un **dispositivo programmabile** per la gestione dell'informazione rappresentata (cioè del **mondo 3**)

Informazione e supporti

**l'informazione è sempre portata da o
trasmessa su o memorizzata in o
contenuta in qualcosa che**

NON E' INFORMAZIONE,

quindi:

***l'informazione è disponibile solo se è accessibile
il supporto (quel qualcosa) che la mantiene***

Informazione / supporti

ogni supporto ha le proprie caratteristiche:

- **alcuni sono migliori per la trasmissione di informazione (es. L'aria, un cavo elettrico),**
- **altri per la sua memorizzazione (es. un foglio di carta, CD)**

La stessa informazione può essere scritta su supporti differenti e uno stesso supporto può portare informazioni differenti

informazione e linguaggio

codifica e decodifica

La relazione tra informazione e supporti è complessa

l'insieme delle regole di tale relazione è una delle basi di un linguaggio

- ◆ la **codifica** è l'operazione di trasmissione/fissazione dell'informazione su un supporto
- ◆ la **decodifica** è il recupero/la lettura di informazione da un supporto

informazione e linguaggio

codifica e decodifica

I singoli atomi di informazione (**simboli**), combinati secondo una configurazione ordinata basata su una codifica/decodifica comprensibile, costituiscono **un messaggio**.

Ovviamente, **più è ampio l'insieme di simboli usato, maggiore l'informazione di quel simbolo:**

se si riduce l'unità elementare a solo due stati/simboli (**sì o no, 1 o 0, acceso o spento...**) si ottiene il massimo sia nell'economia dei simboli usati che nei messaggi possibili, semplificando la comprensibilità

informazione e linguaggio

codifica e decodifica

La **codifica binaria** rappresenta allora **l'alfabeto più semplice possibile** per la codifica di dati e istruzioni:
2 simboli per esprimere tutto.

I computer **si basano su questa codifica**, espressa in **bit** (da *binary digit*, numero binario) combinati in successioni ben formate di **2^n** possibilità, dove **n** rappresenta il numero di bit

2^3 = 8 successioni composte di 3 bit:

000, 001, 010, 100, 100, 101, 110, 111

A una successione composta da 8 bit (**2^8** bit = 256 successioni) si dà il nome di **byte**, da cui i multipli kilobyte (1024 byte), megabyte (1048576 bytes), terabyte, etc.

informazione e linguaggio codifica e decodifica

All'interno delle **apparecchiature digitali** l'informazione è rappresentata mediante livelli di tensione o mediante magnetizzazione di dispositivi appropriati.

Le richieste di affidabilità ci impongono che tali simboli **per una maggiore efficienza siano due:**

si hanno appunto **solo 0 e 1**, corrispondenti a 2 livelli di tensione (standard TTL: 0/5 V; standard RS-232: +12 / -12 V) che vanno a formare la **numerazione binaria**

dai bit alle risorse digitali: la convergenza al digitale

attraverso codifiche standardizzate si **associano alle successioni di bit** valori numerici, caratteri alfanumerici, stati del sistema, valori fisici, etc., per cui si parla di **informazione digitale** (basata su digits, numeri) e di **convergenza al digitale**

la tendenza attuale è cioè a **campionare in digitale tutti i messaggi possibili**

(**testi, numeri, suono, video, azioni meccaniche, onde radio, etc.**) sono quindi espressi tramite la stessa codificazione, quella digitale.

dai bit alle risorse digitali: la convergenza al digitale

Ma che significa **campionare**?

Nella *teoria dei segnali* il campionamento è una tecnica che consiste nel **convertire un segnale continuo nel tempo in un segnale discreto**, valutandone l'ampiezza a intervalli di tempo regolari (nell'audio)

In *statistica* il campionamento è un metodo per **selezionare un campione rappresentativo di un determinato insieme di dati**, in modo da ottenere **risultati generalizzabili**

Più in generale, la conversione da analogico a digitale obbliga a campionare, nel tempo o nello spazio, i dati necessari alla rappresentazione tramite codifica binaria

Analogico e digitale

Analogico

si intende un segnale che è la rappresentazione o trasformazione di una grandezza fisica tramite una sua analoga. Esempi:

secondi (tempo) $\leftrightarrow$ angolo della lancetta dell'orologio

segnale acustico $\rightarrow$ segnale elettrico (microfono)

segnale elettrico $\rightarrow$ segnale acustico (altoparlante)

temperatura $\leftrightarrow$ altezza in mm del termometro a mercurio

In **elettronica**, per analogico si intende il modo di rappresentare il segnale elettrico all'interno di una data apparecchiatura (che lavora sotto potenziale elettrico); il segnale è detto analogico quando **i valori utili che lo rappresentano sono continui (infiniti).**

dai bit alle risorse digitali

le successioni informative binarie, sia che siano disponibili localmente (cioè nella nostra macchina) oppure solo in rete, sono dette **risorse digitali**:

risorse per la funzione informativa, di fonte che rivestono

digitali per il loro formato di codifica (e dunque per la conseguente necessità di disporre di dispositivi di decodifica per accedervi)

dai bit alle risorse digitali

dato / informazione / conoscenza

dato

elemento o serie di elementi conosciuti o verificati; elemento di un'informazione costituito da simboli (De Mauro)

informazione

elemento o dato che permette di venire a conoscenza di qcs.; dato o insieme di dati codificati e immessi in un sistema informatico (De Mauro)

conoscenza

consapevolezza e comprensione di fatti, verità o informazioni ottenuti attraverso l'esperienza o l'apprendimento (a posteriori), ovvero tramite l'introspezione (a priori). La conoscenza è l'autocoscienza del possesso di informazioni connesse tra di loro, le quali, prese singolarmente, hanno un valore e un'utilità inferiori (Wikipedia)

dai bit alle risorse digitali

risorsa

Mezzo o capacità disponibile, consistente in una riserva materiale o spirituale, o in un'attitudine a reagire adeguatamente alle difficoltà (r. economiche, r. naturali, r. intellettuali, r. dell'ingegno, dell'immaginazione, della fantasia) [dal francese *ressource*] – (Devoto-Oli)

Risorse è un termine generico ed è usato in contesti diversi: si parla di risorse finanziarie, risorse idriche, risorse energetiche, risorse bibliografiche, perfino di risorse umane; *Windows 95* (e successivi) prevede l'operazione *Gestione risorse*.

dai bit alle risorse digitali

dato / informazione / conoscenza

conoscenza

- il conoscere qualcosa, il possedere una nozione o un insieme di nozioni organizzate, l'avere una cognizione; l'attività del conoscere mediante lo studio e l'indagine e il suo risultato; facoltà di acquisire cognizioni propria della mente umana (De Mauro)
- *Information that has been comprehended and evaluated in the light of experience and incorporated into the knower's intellectual understanding of the subject . (ODLIS)*

dai bit alle risorse digitali

digitale

Data recorded or transmitted as discrete, discontinuous voltage pulses represented by the binary digits 0 and 1, called bits. In digitized text, each alphanumeric character is represented by a specific 8-bit sequence called a byte. (ODLIS)

In elettronica e in informatica, di sistema o dispositivo che si serve di cifre numeriche per rappresentare dati e grandezze o per riprodurre impulsi fisici (De Mauro)

digital object

In the technical sense, a type of data structure consisting of digital content, a unique identifier for the content, and other data about the content, for example, rights metadata. (ODLIS)

dai bit alle risorse digitali

La risorsa elettronica, sempre nell'ottica della sua catalogazione, **si caratterizza anche per la sua *fisicità*, cioè per la sua disponibilità o indisponibilità *in loco***, ovvero per essere ***ad accesso locale*** o ***ad accesso diretto***

oppure

ad accesso remoto o ***ad accesso a distanza***

Attenzione! Accesso *remoto* identifica il trattamento informativo che facciamo della risorsa, non implica necessariamente il significato consueto di ***distanza fisica***. I dati potrebbero essere nella stanza accanto oppure in Australia, infatti.

dai bit alle risorse digitali(zzate)

Le risorse digitalizzate

Sono, per semplicità, le **riproduzioni digitali di documenti analogici** (su carta, pergamena, pietra, cera, etc.), le rimasterizzazioni digitali di registrazioni audio o video su nastro magnetico.

Una fotografia digitale è una risorsa digitale, una fotografia stampata e scannerizzata è una risorsa digitalizzata; Non sono le ridigitazioni di testi con strumenti elettronici (sono due risorse distinte, quella analogica – il *manoscritto* - e quella digitale). La fotografia apre però il problema del diritto d'autore di cui si dirà più avanti: una foto artistica o professionale di una epigrafe è una risorsa digitale e una risorsa digitalizzata al tempo stesso.

La convergenza digitale

Alberto Salarelli (A. Salarelli – A.M.Tammaro, *La biblioteca digitale*) introduce alcune caratteristiche dei documenti digitali, partendo dall'idea che

Il principale vantaggio della rappresentazione digitale risiede nell'universalità della rappresentazione stessa.

Tutti medium (testo immagine o suono) sono codificati in un'unica forma – sequenza di bit – e tutte le trasformazioni sono per definizioni senza “imprecisioni” a differenza di quelle analogiche che introducono distorsioni e rumore

La convergenza digitale

alcune caratteristiche dei documenti digitali:

Flessibilità I documenti digitali **sono astratti**, in quanto rappresentazioni numeriche, ma per accedervi abbiamo bisogno di interfacce per la decodifica delle sequenze di bit da digitale ad analogico, **da meccanico a sensoriale**. I documenti digitali possono però essere facilmente *combinati*, unendo, mescolando, attaccando (*cut, copy and paste*). **What you store is not what you get**

Simulazione Tali potenzialità di “sforbiciamento” possono non intaccare il documento digitale originario, permettendo di **simulare manipolazioni e ricombinazioni** che i documenti analogici non permettono. La simulazione digitale è sempre reversibile. Il **rapporto originale/copia** è meno significativo che nell'analogico: non c'è differenza tra antografo e apografo, la copia sostituisce l'originale.

Riproducibilità e Conservazione I due termini sono strettamente connessi: “la capacità di un documento elettronico di sopravvivere nel tempo è direttamente proporzionale alla sua facile riproducibilità. E inversamente proporzionale alla labilità dei supporti”. **L'opposto rispetto ai documenti manoscritti e non molto diverso dai documenti a stampa.**

I formati delle risorse digitali

I formati dei dati in formato digitale sono la **modalità con cui i dati vengono rappresentati elettronicamente in modo che i programmi possano elaborarli.**

Per elaborare un file, un programma deve sapere come i dati vi sono “memorizzati”

Il formato specifica la codifica adottata, ovvero la **corrispondenza fra la rappresentazione binaria e i dati rappresentati**

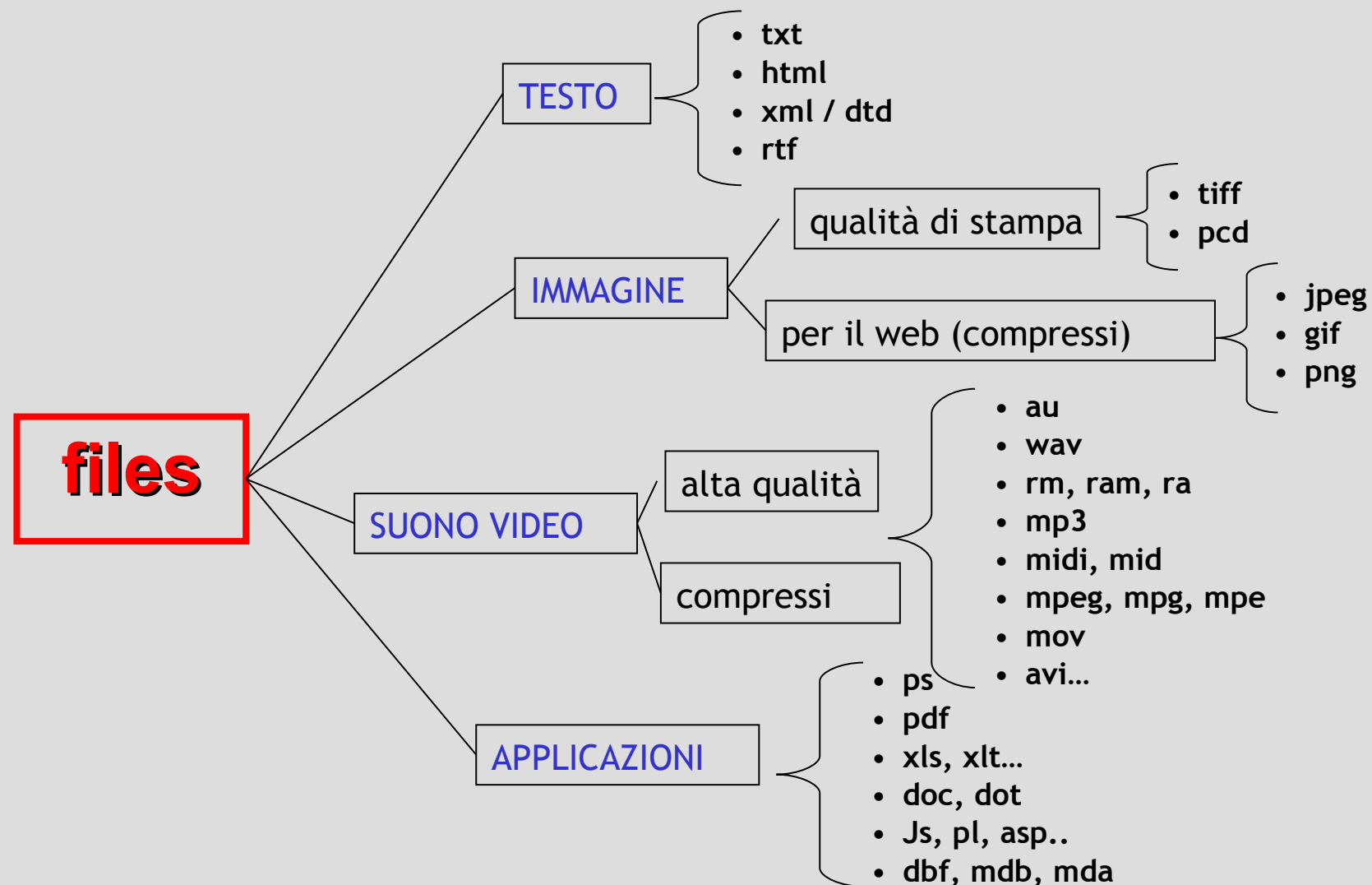
I formati delle risorse digitali

Per quanto riguarda i formati dei **file** (*collection of computer data which is stored under one filename*) è necessario distinguere **tra la struttura logica del dato** (elementi unitari minimi) **e quella fisica** (che cambia a seconda delle conformazione hardware e software etc.).

Poi, i file sono raggruppati secondo lo **standard MIME**, nato soprattutto in seguito alla diffusione del Web e dei browser (che devono associare facilmente i file in rete coll'applicazione che serve a decodificarli e renderli accessibili). MIME divide i files in MIME/types:

TEXT, IMAGE, AUDIO, VIDEO, APPLICATION

I formati delle risorse digitali



I formati delle risorse digitali

Type:text

(dati codificati con combinazioni di mappe caratteri - charset - ed eventuali formattazioni)

Subtype and extensions:

plain	txt, text	Cfr. RFC 2046
html	htm, html	
sgml	sgml	http://www.w3.org/MarkUp/SGML/
xml	xml, dtd, x	http://www.w3.org/MarkUp/
rtf	rtf	www.soft-land.org/documenti/rtf.html

I formati delle risorse digitali

Type:image (bit-mapped / vettoriali)

Subtype and extensions:

tiff	tiff, tif	bit-mapped -paint
gif	gif	bit-mapped -paint (http://www.auditmypc.com/acronym/GIF.asp)
jpeg	jpg, jpeg, jpe, jfif, pjpeg, pjp	bit-mapped -paint (http://www.jpeg.org)
x-photo-cd	pcd	bit-mapped -paint
png	png	bit-mapped -paint
x-ms-bmp	bmp	bit-mapped -paint
fractals image format	fif	Vettoriale -draw (http://www.auditmypc.com/acronym/FIF.asp)

I formati delle risorse digitali

Type: audio

(ha bisogno di un apparecchiatura di output audio e può essere digital-born o riversato da analogico. La compressione elemento fondamentale)

Subtype and extensions:

basic	au, snd	compresso di base
x-wav	wav	compresso
x-pn-realaudio	rm, ra, ram	ideale per lo streaming
mpeg	mp2, mp3, mpa, abs, mpega	mp3 miglior rapporto compressione/prestazioni
midi	midi, mid	non è la registrazione di un suono ma i comandi per un brano musicale

I formati delle risorse digitali

Type: video

(immagini in movimento: digital-born o riversate da analogico. La compressione elemento fondamentale)

Subtype and extensions:

mpeg	mpeg, mpg, mpe, vbs...	
mpeg2	mpv2	in uso per i dvd free (iso/eic-13818)
quicktime	qt, mov, moov	proprietario, con funzioni speciali
x-msvideo	avi	
x-sgi-movie	movie	

I formati delle risorse digitali

Type: application (files di tutti i tipi precedenti ma codificati con programmi proprietari, per la cui codifica è necessario un software specifico)

Subtype and extensions:

postscript	ai, eps, ps	testi/immagini ad alta formattazione per la stampa
pdf	pdf	idem
msword	doc, dot	testo formattato
vnd.ms-excel	xls, slt, xlm, xld...	fogli elettronici
vnd.musician	mus	spartiti musicali
vnd.ms-access	mdb, mda, mde	database
x-javascript	js, mocha	software eseguibile
x-perl	pl	idem

I formati delle risorse digitali

I formati digitali si possono classificare anche a seconda se sono:

- **proprietary** posseduti in esclusiva da un soggetto che ne mantiene segreto il funzionamento
- **aperti** sono resi pubblici e sono documentati
- **liberi** si lascia libertà di utilizzare, studiare, modificare e redistribuire i formati, a patto di non sfruttarli commercialmente (GNU-GPL)

I formati delle risorse digitali

formati proprietari / formati aperti

[..] una tecnologia si dice “**aperta**” quando è resa pubblica ed è documentata esaustivamente.

“Aperta” si oppone a “**proprietaria**”, che indica una tecnologia posseduta in esclusiva da un soggetto che ne mantiene segreto il funzionamento e può modificarla a proprio piacimento

Al requisito della pubblicità per alcuni si aggiunge anche quello relativo alla proprietà

*dall' Indagine conoscitiva sul software a codice sorgente aperto nella Pubblica Amministrazione -
Rapporto della Commissione*

I formati delle risorse digitali

formati liberi

- Libertà di **utilizzare** il formato per qualunque scopo;
- Libertà di **studiare** il formato ed adattarlo alle proprie necessità. Conoscerne le specifiche è un prerequisito;
- Libertà di **redistribuire** le specifiche del formato;
- Libertà di **modificare** le specifiche e redistribuirne le versioni modificate.

I diritti di proprietà intellettuale

Accesso aperto – Open access

Movimento internazionale, in ambito di letteratura scientifica, per l'accesso alla produzione scientifica. Alla sua base la **Dichiarazione di Berlino** (2003):

[<http://www.zim.mpg.de/openaccess-berlin/ BerlinDeclaration_it.pdf>](http://www.zim.mpg.de/openaccess-berlin/BerlinDeclaration_it.pdf)

- l'informazione deve essere resa largamente e prontamente disponibile alla società
- si devono sostenere nuove possibilità di disseminazione della conoscenza attraverso **l'accesso aperto**: una fonte estesa del sapere umano e del patrimonio culturale che siano stati validati dalla comunità scientifica
- il Web deve essere sostenibile, interattivo e trasparente
- i contenuti e i mezzi di fruizione compatibili e ad accesso aperto

I diritti di proprietà intellettuale

Creative Commons <http://www.creativecommons.it/>

Un set ormai diffuso di licenze standard è l'iniziativa **Creative Commons**, nata nel 2001, che si collega al modello Open Access. Ha rilasciato un set di **sei licenze standard** a livello internazionale che offrono uno spettro di opzioni intermedie fra il regime *full copyright* (tutti i diritti riservati) e il *public domain* (dominio pubblico).

Le licenze Creative Commons combinano fra loro **quattro diversi elementi base**: *Attribution* (Attribuzione), *Non-Commercial* (Non commerciale), *Non-derivative* (Non opere derivate), *Share-alike* (Condividi allo stesso modo).

Ogni licenza è disponibile in tre versioni: per i legali; per gli utenti; in formato leggibile dalla macchina (in XML-RDF).

I diritti di proprietà intellettuale

Le licenze Creative Commons **consentono:**



- la visualizzazione pubblica; la distribuzione; la copia; l'integrazione dell'opera in una o più opere collettive, obbligando sempre e comunque a citare la fonte del materiale riutilizzato.

Fra le altre possibili **restrizioni e condizioni:**

- l'utilizzazione per scopi non commerciali; la limitazione del riuso per opere derivate.

Prima di poter accedere a un qualsiasi contenuto, l'utente dovrà esplicitamente accettare i termini e le condizioni d'uso. Le licenze Creative Commons sono state utilizzate dall'OAI Open Archives Initiative e dall'iniziativa Dublin Core (di ci si dirà parlando di metadati descrittivi).

I diritti di proprietà intellettuale

L'espressione dei **Termini di disponibilità** o delle **Regole commerciali** per le risorse è demandata sulla Rete ai **metadati per la gestione dei diritti**. Per chiarezza sia degli utenti sia dei fornitori di servizi si raccomanda di distinguere i metadati relativi ai diritti di proprietà intellettuale dalla risorsa.

Per esprimere i termini delle licenze senza ambiguità negli schemi di metadati amministrativi si possono usare gli standard REL (Rights Expression Language). Lo schema ODRL (Open Digital Rights Language) è un'applicazione open source di REL in XML; XrML è invece uno standard proprietario.

I progetti dovrebbero mantenere (ovvero gestire, aggiornandoli) i dati sui diritti che detengono o acquisiscono in modo coerente, così da poterli condividere con sistemi esterni in un formato standard.

I diritti di proprietà intellettuale

Digital Rights Management - DRM

sono i sistemi tecnologici mediante i quali i titolari di diritto d'autore (e dei diritti connessi, vedi gli editori) possono esercitarli in digitale rendendo protette, identificabili e tracciabili le opere.

Tramite i DRM, i file (testo, immagine audio, video, programmi) vengono codificati e criptati in modo da garantire una più difficile diffusione, impedimenti all'utenza e consentirne un utilizzo:

- **limitato** (ad esempio solo per determinati periodi di tempo o per determinate destinazioni d'uso);
- **predefinito** nell'uso in base alla licenza d'accesso fornita (separatamente) agli utenti finali.

I diritti di proprietà intellettuale

Anche per i **file audio** regolarmente acquistati sul Web sono stati integrati schemi di DRM per costringere l'utente ad una serie di limitazioni, come ad esempio il numero o i tipi di dispositivi.

Molti produttori di **eBook** usano una implementazione simile di DRM per limitare il numero di computer sui quali l'eBook può essere visualizzato o anche quante volte può essere letto. Il **Social DRM** è un metodo che non limita l'utilizzo del file, ma agisce inserendo al suo interno delle informazioni sul proprietario difficili da rimuovere: si personalizza e rende unico l'eBook scaricato con il nome e il cognome di chi l'ha acquistato e con alcune informazioni sull'acquisto (data e numero della transazione, per esempio).

L'**Adobe DRM** è un sistema di protezione che controlla l'identità dell'utente, contatta una finanziaria per concordare il pagamento, decodifica il file, e gli assegna una chiave, una specie di password, per un accesso futuro.

I diritti di proprietà intellettuale

DRM tramite marcatura e impronte digitali

I sistemi multimediali possono considerare la possibilità di apporre una *filigrana* o un'impronta digitale sui materiali che producono allo scopo di salvaguardare i diritti su di essi detenuti.

La filigrana digitale, o marchio elettronico (*watermarking*), è un segnale (marchio) incorporato permanentemente nei dati digitali, ovviamente apposto informando dettagliatamente l'utenza della sua esistenza e delle sue caratteristiche.

A seconda della specifica implementazione, il watermark può essere:

- ben visibile, come il logo di una emittente televisiva;
- nascosto, tramite *steganografia* (in questo caso è più difficile da rimuovere).

I supporti per le risorse digitali

Premessa

- Le risorse digitali da conservare sono *oggetti per musei* oppure *oggetti per archivi*?
- Nelle procedure di riproduzione si può scegliere di perdere parte dei contenuti, pur conservando l'efficacia della comunicazione

Quindi:

Non possiamo conservare altro che quello che scegliamo di conservare

I supporti per le risorse digitali

I supporti di memorizzazione

- Si possono individuare tre grandi famiglie tecnologiche di supporti di memorizzazione:
 - I dischi rigidi
 - I supporti mobili magnetici
 - I supporti mobili ottici
 - I supporti flash

I supporti per le risorse digitali

I dischi rigidi

I dischi rigidi (Hard Disk o HD) non sono supporti removibili, anche se sul mercato sono disponibili dei dischi esterni, da usare come memoria di massa aggiuntiva.

La tecnologia è magnetica, e si sono usati per i PC dal 1985 circa.

Un Hard Disk e' composto da due o piu' dischi con entrambe le superfici ricoperte da una sostanza magnetica che consente la memorizzazione dei dati. Su ogni superficie sono presenti due testine: una di lettura ed una di scrittura, che leggono o modificano i dati presenti su quella superficie. I dischi sono disposti uno sopra l'altro e ruotano quindi su di un asse comune.

I supporti per le risorse digitali

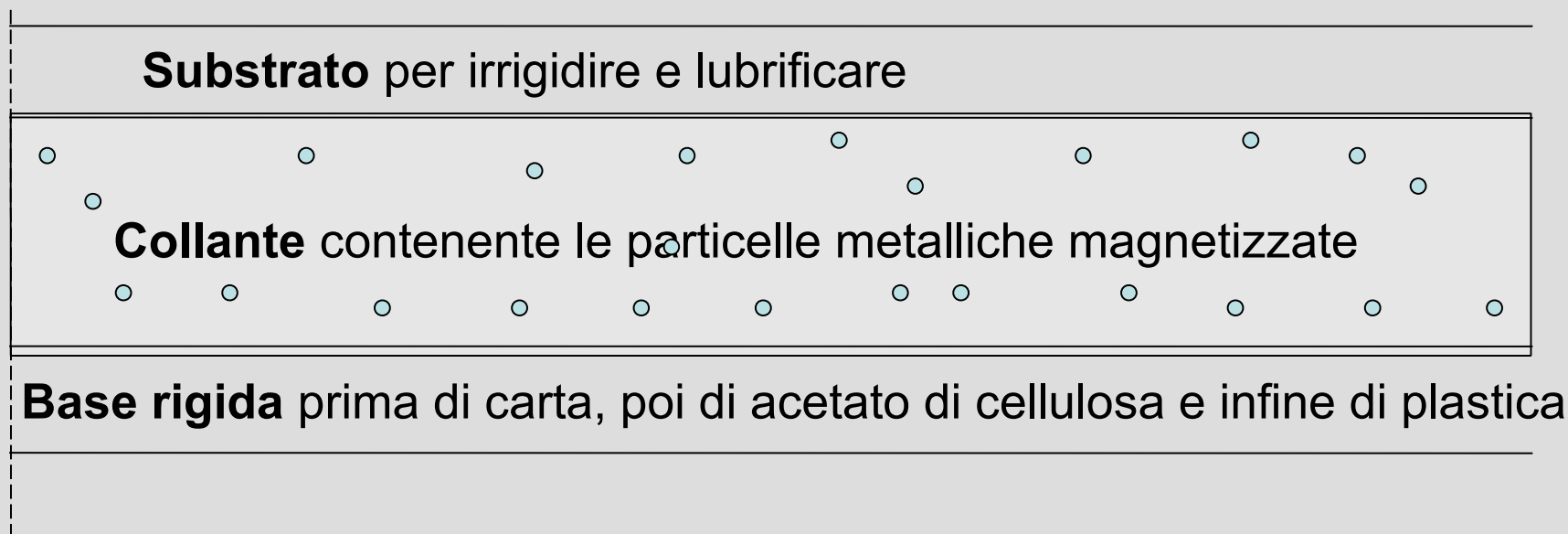
I supporti magnetici

- Nascono nel corso della II guerra mondiale a sostituire i fili magnetici.
- Si basano sulla carica magnetica di particelle metalliche, letta da una testina.
- Sono supporti sequenziali e non hanno alcuna struttura al loro interno che rechi informazione rispetto ai file contenuti.
- Sono inseriti in strutture protettive (cartridge) o aperti, sono anelli chiusi oppure nastri con un inizio e una fine

I supporti per le risorse digitali

I supporti magnetici

Il nastro è composto di vari strati:



Oggi esistono cartridge molto evolute e velocissime (DAT), con chip che controllano l'integrità dei dati e li "firmano".

I supporti per le risorse digitali

I supporti magnetici – rischi per la conservazione

- ossidazione delle particelle magnetiche
- Smagnetizzazione per la vicinanza di fonti magnetiche (cellulari, trasformatori, etc.)
- perdita del lubrificante
- degradazione del collante
- deformazioni meccaniche dovute a trazione eccessiva, al calore etc.

I supporti per le risorse digitali

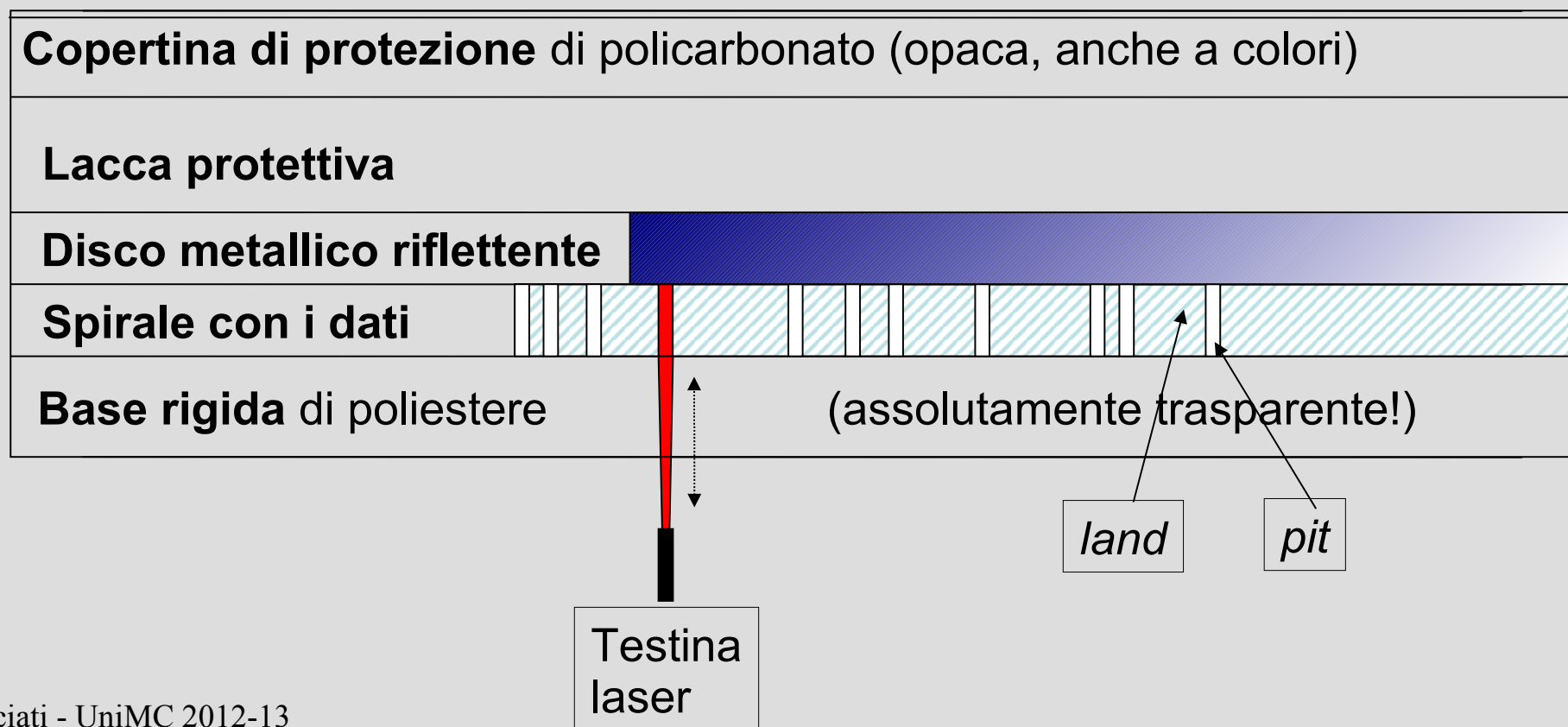
I supporti ottici

- Idea del 1958 ma brevetto del 1969, per la registrazione di audio analogico, cercando di impedire l'attrito tra testine e dischi
- la testina è una fonte di luce laser
- tre tipi essenziali (vale sia per CD che per DVD):
 - **ROM (Read Only Memory)**
 - **WORM (Write Once Read Many)**
 - **RW (Read Write)**

I supporti per le risorse digitali

I supporti ottici

Il disco è composto di diversi strati, che per il DVD sono di più e per i dischi double-face sono raddoppiati:



I supporti per le risorse digitali

I supporti ottici

- rischi per la conservazione:
 - graffi e lesioni sulla base trasparente
 - lesioni da vernici acide che attraversano la base trasparente
 - ossidazione del disco metallico
 - polvere (modifica chimicamente il policarbonato e crea microfratture)
 - stabilità della forma, che deve essere perfettamente in piano

I supporti per le risorse digitali

I supporti flash

La **memoria flash** è una memoria basata su mini-transistor caricati con elettroni, che per le sue prestazioni può anche essere usata come memoria a lettura-scrittura.

Trattandosi di memoria a stato solido, non presenta alcuna parte mobile quindi è piuttosto resistente alle sollecitazioni e agli urti, inoltre è estremamente leggera e di dimensioni ridotte. È infatti molto usata nelle **fotocamere digitali**, nei **lettori di musica portatili**, nei **cellulari**, nei **pendrive** (chiavette USB), nei **palmari**, nei *netbook* e in molti altri dispositivi che richiedono un'elevata portabilità e una buona capacità di memoria per il salvataggio dei dati.

I supporti per le risorse digitali

I supporti flash

I *Solid State Drive* sono un prodotto già da tempo presente sul mercato, inizialmente dedicato ad applicazioni particolarmente critiche di tipo militare e industriale. Oggi si propone invece come sostituto per hard disk di portatili o altri dispositivi portatili. Le prestazioni rispetto un hard disk tradizionale sono superiori: maggiore velocità nella lettura e scrittura dei dati, maggiore affidabilità, altissima resistenza agli shock meccanici, bassissimi consumi.

I dischi allo stato solido non consentirebbero, come i supporti magnetici, un numero illimitato di scritture; tuttavia all'interno dei dispositivi di archiviazione sono inseriti, a livello hardware, degli speciali algoritmi che evitano la località dell'accesso alle celle di memoria prolungando notevolmente la vita del dispositivo.

Le risorse digitali aggregate

le risorse devono essere associate in insiemi/collezioni che siano organizzati / associati, per avere la garanzia di essere correttamente contestualizzate .

Semplificando, i sistemi di risorse (collezioni) possono essere:

- **strutturati in Data Base**, quindi rispondere a una logica e sintassi definita, che può essere anche molto complessa (in certe condizioni sono detti anche **sistemi informativi**)
- **Associati tramite metadati strutturali** (ne parleremo più avanti)
- **basarsi su collegamenti ipertestuali** (siti e applicazioni web, cui dedichiamo il modulo 2)

Le risorse digitali aggregate

database

un **database** è una **raccolta di dati permanenti**, gestita da un **elaboratore** e suddivisa in

Metadati o schema: definizioni che ci dicono quali regole valgono sui dati, quali valori possono essere validi (vincoli di integrità), come i dati sono strutturati e collegati tra loro. Lo schema può cambiare nel tempo, non ha alcun vincolo con i programmi che accedono al database e deve essere definito prima dei dati.

Dati: rappresentazione (astrazione) dei fatti, ma può anche includere previsioni, supposizioni ed ipotesi conformi ai Metadati (o schema).

I dati sono organizzati in **insiemi omogenei** (stessa struttura) e sono collegabili con altri insiemi. Il **modello dei dati** è un insieme di astrazioni che descrivono le associazioni tra i dati.

Le risorse digitali aggregate

Dbms, sql

Data Base Management System. Sistema centralizzato (programmi coordinati) o distribuito (rete) che permette di memorizzare, modificare ed estrarre informazioni da un database. **Garantisce l'integrità dei dati** (unica raccolta di dati anziché copie distinte scoordinate che potrebbero causare duplicazioni, ridondanze). **Organizza le informazioni** del database secondo la struttura di un database gerarchico, di un database di rete o di un database relazionale. **Garantisce l'accesso** alle informazioni solo a persone autorizzate, p.e. tramite una password. Un DBMS oltre a stabilire schemi organizzativi e di controllo, rende le informazioni accessibili agli utenti, tramite query. Esempi di DBMS: Oracle, Microsoft SQLServer, Basis, MySQL, PostgreSQL, InterBase.

Structured Query Language, linguaggio di interrogazione strutturato. SQL, oltre a creare tabelle bidimensionali permette la formulazione di query (interrogazioni) sulle relazioni (tabelle). SQL è il linguaggio più utilizzato per definire database.

Le risorse digitali aggregate

Tipi di database

Le basi di dati possono avere **varie strutture**, tipicamente, in ordine cronologico:

1. **gerarchica** (rappresentabile tramite un albero - anni sessanta),
2. **reticolare** (rappresentabile tramite un grafo - anni sessanta),
3. **relazionale** (attualmente **il più diffuso**, rappresentabile mediante tabelle e relazioni tra esse - anni settanta),
4. **ad oggetti** (estensione alle basi di dati del paradigma "Object Oriented", tipico della programmazione a oggetti - anni ottanta),
5. **semantica** (rappresentabile con un grafo relazionale - inizio anni duemila).

Un requisito importante di una buona base dati consiste nel *non duplicare inutilmente le informazioni in essa contenute*: questo è reso possibile dai gestori di database relazionali, che consentono di salvare i dati in tabelle che possono essere collegate.

Le risorse digitali aggregate

data base relazionale

L'assunto fondamentale del modello relazionale è che **tutti i dati sono rappresentati come relazioni**; i dati sono manipolati con gli operatori dell'algebra relazionale.

Il modello relazionale consente al progettista di database di creare una **rappresentazione consistente e logica dell'informazione**.

La **consistenza** viene ottenuta inserendo nel progetto del database appropriati **vincoli**, normalmente chiamati **schema logico**.

La struttura base del modello relazionale è il **dominio** o **tipo di dato**, definito come l'insieme dei valori che può assumere un determinato attributo.

Un **attributo** è una coppia ordinata di "nome di attributo" e "nome di tipo", mentre un valore di attributo è un valore specifico valido per quel tipo di dato. Una **relazione** consiste di una testata e di un corpo, dove la testata è un insieme di attributi e il corpo è un insieme di n valori possibili. La **tabella** è la rappresentazione grafica normalmente accettata per rappresentare la relazione.

Le risorse digitali aggregate

CMS, blog

Blog e forum sono particolari **Content Management System (CMS)**, implementati da database (realizzati ad esempio con DBMS come MySQL). I database vengono interrogati e gestiti generalmente attraverso istruzioni PHP, VBScript (ASP) o Perl. L'insieme dei file che contengono le istruzioni per operare con la base di dati e per impaginare gli interventi, costituisce il pacchetto di blogging installato nel server web.

Blog: contrazione di Web e Log, commenti (“tracce”) sul web. Particolare pagina web strutturata come un diario che raccoglie testi aperti ai commenti dei lettori. Gli utenti possono scrivere nelle pagine-blog pensieri, opinioni, poesie, esperienze, oppure inserire dei link su specifici argomenti, nonché immagini. Ogni nuovo intervento inviato dall'utente viene inserito in cima alla pagina web, riportandone la data e l'ora dell'inserimento. L'aspetto (layout) delle pagine-blog può essere modificato utilizzando specifici **modelli grafici (Template)**.

Le risorse digitali aggregate

Possiamo distinguere, semplificando e solo a scopo didattico, **due tipi sostanziali di banche dati**, rispetto alla organizzazione delle macchine su cui sono conservati i dati e dove si trova il software gestionale:

- con **dati organici** al DBMS
- con **dati indipendenti** dal DBMS

In generale, un **database distribuito** è un database che si trova sotto il controllo di un DBMS nel quale gli archivi *non sono memorizzati sullo stesso computer*.

Il database può essere dislocato in più computer situati nello stesso luogo, oppure distribuito in una rete di computer connessi tra loro.

Le risorse digitali aggregate

- **con dati organici al DBMS**

in questo caso la struttura, il formato, le maschere di inserimento, le query e le modalità di aggregazione, ordinamento e output (anche Web) sono interamente gestite dal DBMS.

Questo se semplifica la fase di progettazione iniziale (“chiavi in mano”) può creare delle limitazioni e difficoltà per la portabilità e l’interoperabilità dei dati.

Le risorse digitali aggregate

- **con dati indipendenti dal DBMS**

in questo caso

- la struttura dei dati sono basati su formati leggeri e standard (XML prima di tutto)
- le maschere di inserimento possono essere costruite con software di editing anche molto semplici e gratuiti
- le query e le modalità di aggregazione, ordinamento, sicurezza, conservazione sono gestite dal DBMS
- l'output e l'interoperabilità possono essere indipendenti dal DBMS, "tornando" ai dati originari

Le risorse digitali aggregate

i sistemi informativi

attualmente, nell'approccio alle risorse digitali per Archivi e Biblioteche, soprattutto, ma anche per le risorse culturali storico-artistiche, si finanziano progetti che comprendano le funzioni di **descrizione, aggregazione, amministrazione, gestione, conservazione e comunicazione**: i SI

“Il sistema informativo è costituito dall'insieme delle informazioni utilizzate, prodotte e trasformate da un'azienda durante l'esecuzione dei processi aziendali, dalle modalità in cui esse sono gestite e dalle risorse, sia umane, sia tecnologiche, coinvolte.”
(Wikipedia)

Le risorse digitali aggregate

Se dal punto di vista dei formati descrittivi associativi e gestionali valgono tutte le cose dette finora, vale la pena di accennare qualcosa sulle possibili **architetture dei sistemi informativi**.

Sinteticamente, questi possono essere **distribuiti o centralizzati**, ma i modelli sono in realtà ben più complessi, anche in relazione alla “dimensione”, cioè alla complessità e quantità dei dati e al contesto organizzativo.

Le risorse digitali aggregate

i **sistemi centralizzati** raccolgono i database (anche se magari non le funzioni gestionali e quelle di output Web) in un sistema centrale, alimentato tramite connessione di rete da n terminali di immissione periferici.

Non cambia molto se questa immissione avviene tramite poli intermedi, anche strutturati a più livelli.

Le risorse digitali aggregate

i pro e i contro dei **sistemi centralizzati**

- **i pro**: facilitano la omogeneità formale e la validazione dei dati e la creazione di liste d'autorità, evitando ripetizioni e sovrapposizioni (pensiamo al modello SBN)
- **i contro**: l'immissione dei dati è spesso lenta e faticosa, e non è sempre agevole valorizzare i contesti locali

Le risorse digitali aggregate

i sistemi distribuiti

si basano sulla costruzione di n sistemi locali autonomi, che riversano tutti o parte dei loro dati su un sistema centrale deputato a funzioni particolari, tra cui quella Web. I sistemi locali sono autonomi ma nel loro rapporto col “centro” sono allineati attraverso procedure automatiche (ad es. OAIS) oppure no.

Le risorse digitali aggregate

i pro e i contro dei **sistemi distribuiti**

- **i pro**: valorizzano i contesti locali permettendo personalizzazioni e applicazioni specifiche; velocizzano immissione e gestione dei dati
- **i contro**: le procedure di omogeneizzazione e validazione dei dati sono più delicate

Le risorse digitali aggregate

Collezioni digitali

Nei progetti di digitalizzazione di materiali analogici, poi, in seguito a un processo di selezione e di digitalizzazione si creano **collezioni digitali**, cioè aggregazioni ex novo di oggetti digitali associati sulla base di criteri diversi: associazione originaria (serie archivistiche o bibliografiche, collezioni artistiche, pagine di un libro etc.), tema, cronologia, obiettivi formativi o didattici, esposizioni, dimostrazioni, etc.

L'associazione, come detto, può avvenire via *dbms*, *ipertesto per il web*, *metadati strutturali*. Ne parleremo ancora trattando di progetti.

Le risorse digitali aggregate

Collezioni digitali

Si capirà fin da ora come siano importanti le informazioni su **chi** ha effettuato l'aggregazione, sulla base di **che criteri**, **quando**, **per conto di chi**, adottando **che tecniche**, etc.

Le collezioni meritano di essere valorizzate anche oltre l'effimero, specie se finalizzate ad esser **pubblicate sul web**. Devono insomma essere **identificate, recuperate, scelte, indicizzate** come collezioni, non solo attraverso le singole risorse che le compongono!

Le risorse digitali aggregate

Descrizione delle collezioni

Per descrivere le collezioni sui MICHAEL, ad esempio, si raccolgono informazioni su 5 entità distinte, in relazione tra loro:

- Collezione digitale
- Eventuale collezione/i fisica
- Organizzazione/i coinvolt/e
- Progetto/i all'origine della collezione
- Servizio/i che mettono la collezione a disposizione degli utenti



La descrizione delle collezioni

(sistema relazionale in MICHAEL)

