

# **Tutorial**

ad uso dei progetti di digitalizzazione  
cofinanziati dalla Regione Lombardia

versione 1.2  
(dicembre 2009)

# Sommario

La conversione digitale.....	3
La selezione dei materiali.....	5
Il diritto d'autore .....	8
I file immagine .....	11
Ocr e Codifica.....	18
Il controllo della qualità della digitalizzazione .....	19
Metadati .....	21
Strumenti tecnologici.....	23
Presentazione dei dati.....	26
Conservazione digitale.....	28
Gestione del progetto.....	31
Il personale .....	32
Appendice: Flusso decisionale .....	34

## La conversione digitale

Con le espressioni “conversione digitale”, “acquisizione digitale” o più semplicemente con il termine “digitalizzazione” ci si riferisce a quel processo che trasforma un oggetto analogico di qualsiasi natura (un documento cartaceo, una diapositiva, una videocassetta, ecc.) in un oggetto digitale; tale trasformazione avviene grazie alla produzione, tramite una macchina fotografica digitale, uno scanner o un registratore digitale, di uno o più file che riproducono l’oggetto analogico in questione.

Nel presente tutorial, per quello che riguarda la questioni di carattere tecnico, ci riferiremo alla digitalizzazione di oggetti cartacei (libri, carte geografiche, stampe, ecc).

La scelta da parte di un’istituzione di avviare un progetto di digitalizzazione comporta la necessità di affrontare questioni sia di natura strategico-culturale, che tecniche, ovviamente dopo aver risolto quelle economiche.

Il Manuale Minerva per l’interazione con gli utenti del Web culturale dedica un paragrafo alle possibili definizioni di ‘biblioteca digitale’.

Una biblioteca digitale è una biblioteca in cui le collezioni sono in formato digitale (non sono dunque compresi i formati a stampa, microfilm o altri supporti) e rese accessibili tramite computer. Sia la biblioteca “fisica” che quella digitale offrono un servizio che consente agli utenti, attraverso l’uso di strumenti specifici, di effettuare ricerche nell’ambito delle proprie collezioni.

I contenuti digitali sono di norma accessibili in remoto tramite reti di computer. Una biblioteca digitale è, da un punto di vista strettamente tecnico, un sistema di *information retrieval*. I bibliotecari possono considerare una *digital library* come uno spazio alternativo di mediazione culturale e conversazione, simile ad una biblioteca ma in un contesto digitale.

Le componenti fondamentali di una biblioteca digitale sono:

- la **collezione**, che può essere costituita da testi, immagini, video, suoni e metadati ed essere di carattere permanente o temporaneo, essere stata creata cioè per un lasso di tempo prefissato
- i **servizi di accesso**, che devono permettere agli utenti di trovare rapidamente e facilmente ciò che cercano ed estendere la ricerca ai documenti connessi. Gli strumenti d’accesso includono l’interfaccia utente, i sistemi di ricerca, identificazione, navigazione e connessione delle informazioni desiderate
- l’**utente**, che si muove da solo senza intermediazioni e senza limiti spaziotemporali. Dal momento che l’utente è un soggetto attivo, ne consegue che il documento digitale da lui cercato può essere dinamico e avere un “periodo di vita” differente a seconda degli usi che ne vengono fatti. Le esigenze degli utenti possono essere individuate con l’uso di meccanismi di *feedback* (vedi 2.6).

Allo stato attuale nel Web sono rintracciabili numerose realizzazioni definite *digital library*:

1. *repository* di documenti o pubblicazioni di carattere tematico o accademico accessibili via Web (basati, ad esempio, sulla piattaforma *open source* DSpace)
2. *repository* digitali caratterizzati da una prevalenza di contenuti prodotti nell’ambito di un progetto specifico
3. collezioni di pubblicazioni o materiale multimediale originariamente prodotte in forma analogica, rese accessibili via Web a digitalizzazione avvenuta

4. siti web di biblioteche, archivi, musei e altre istituzioni culturali, che possono essere definiti *digital library* dal momento che offrono documentazione, pubblicazioni e contenuti multimediali.<sup>1</sup>

---

<sup>1</sup> Minerva Working Group. *Manuale per l'interazione con gli utenti del Web culturale*.  
<[http://www.minervaeurope.org/publications/Handbookwebuserinteraction\\_it.pdf](http://www.minervaeurope.org/publications/Handbookwebuserinteraction_it.pdf)>

## *La selezione dei materiali*

In questa sezione vengono forniti alcuni criteri che possono aiutare nella scelta del materiale da digitalizzare.

La scelta del materiale da digitalizzare può essere rivolta

a singoli oggetti che possono o meno condividere alcune caratteristiche comuni, sia di natura materiale che intellettuale

a serie, collezioni o porzioni di raccolte.

Le motivazioni che spingono alla digitalizzazione possono essere diverse; nel caso dei progetti di acquisizione digitale cofinanziati dalla Regione Lombardia, è necessario tenere presente che gli oggetti creati entreranno a far parte della Biblioteca Digitale della Lombardia. Per questo motivo, sarebbe utile che i materiali da acquisire possano essere in qualche modo riconducibili alle strategie culturali evidenziate nel Manifesto della Biblioteca Digitale della Lombardia.

Per evitare possibili duplicazioni, i materiali oggetto di scelta dovranno essere segnalati alla Regione che potrà provvedere a svolgere adeguati controlli; spetta invece alla singola istituzione una preventiva verifica per accertare l'eventuale digitalizzazione del documento presso i più importanti progetti di digitalizzazione.

I materiali dovranno inoltre essere segnalati alla Soprintendenza Beni Librari.

### *Checklist di ausilio alla selezione*

- Il progetto duplica anche solo in parte altre iniziative? Oppure si riferisce a materiali che sono posseduti unicamente dall'istituzione in questione?  
Questa valutazione implica la conoscenza dei progetti di digitalizzazione sia di ambito regionale che nazionale o internazionale. Ad esempio, determinati autori possono essere stati digitalizzati in iniziative di valore nazionale (ad es. il progetto relativo alla collana Scrittori d'Italia all'interno della BDI) o internazionale (ad es. le opere di personaggi italiani che hanno avuto un ruolo di rilievo in altri paesi potrebbero esser state acquisite in progetti esteri).
- I risultati del progetto di digitalizzazione hanno un valore intellettuale duraturo nel tempo?  
Ossia ci deve essere la certezza di una durevole utilità intellettuale anche nel caso in cui il progetto sia legato ad eventi o circostanze contingenti (mostre, ricorrenze, ecc.).
- La digitalizzazione migliorerà la leggibilità dell'originale?  
Valido non solo per i libri antichi, ma per tutti i documenti stampati, ad esempio, con inchiostri acidi o su carta divenuta col tempo molto fragile.
- La digitalizzazione darà valore aggiunto alla collezione?  
La risposta a questa domanda ha numerose possibili implicazioni: ad esempio 1) i documenti in questione attraverso la digitalizzazione acquisteranno modalità di fruizione più accurate? 2) sarà possibile, unificando i risultati di progetti svolti da biblioteche diverse, ottenere la completa acquisizione di un'opera (un periodico, un annale, delle carte geografiche di un territorio, sparsi tra biblioteche diverse)? 3) la consultazione in simultanea sullo schermo del computer di pagine diverse di un medesimo documento permette di effettuare dei raffronti e degli studi impossibili nell'originale cartaceo?.
- La digitalizzazione aumenterà l'utenza dei materiali in oggetto?

La digitalizzazione riguarderà documenti che, a causa o del loro stato di conservazione o della difficile modalità di fruizione o della scarsa conoscenza da parte dell'utenza allargata sono, per adesso, consultati solo da un ristretto numero di lettori, nonostante possano interessare un più ampio pubblico (magari di tipologia differente)?

- Il progetto potrebbe costituire un collegamento tra l'Istituzione e una qualche altra istituzione, in special modo del territorio?

Ossia si potrebbe pensare di dare particolare rilievo a quei progetti che potrebbero costituire un ponte con istituzioni diverse presenti sul territorio. Ad esempio, il progetto di digitalizzazione di materiali di storia municipale potrebbe collegarsi con le attività del museo civico; oppure la digitalizzazione di pubblicazioni di contenuto geografico relative al territorio, sia storiche che recenti, potrebbe risultare utile anche a specifici uffici comunali; l'acquisizione di un fondo librario, o di alcuni esemplari, appartenuti ad una famiglia importante della città (per motivi commerciali, politici, artistici, ecc) potrebbe costituire un collegamento con l'eventuale documentazione archivistica relativa custodita nell'archivio di riferimento.

- Il progetto si inserisce nella Biblioteca Digitale della Lombardia?

Ossia verranno acquisiti materiali che testimoniano la cultura lombarda o che possono contribuire allo studio e alla comprensione della stessa?

#### *Motivazioni per la digitalizzazione*

- Le pubblicazioni in questione sono rare o di difficile reperibilità? Potrebbero essere reintegrate nel caso in cui non esistessero più?
- Le pubblicazioni in questione presentano alcune caratteristiche, annotazioni, firme, ex libris eventi particolari collegati alla creazione o all'esistenza dell'oggetto, che ne aumentano il valore?
- Le pubblicazioni contengono informazioni direttamente relative a persone, posti o eventi collegati all'istituzione che cura la digitalizzazione?
- Le pubblicazioni rispecchiano alcuni aspetti significativi per la comunità che la digitalizzazione metterebbe in risalto?

6

#### *Diritti*

- Gli oggetti da digitalizzare sono liberi da diritti?
- In caso negativo, l'Istituzione possiede i diritti per gli oggetti da digitalizzare o li può ottenere facilmente?

#### *Utenza*

- Quale categoria/e di utenza sarà maggiormente interessata ai prodotti delle digitalizzazioni?
- I materiali sono di interesse ampio o specifico?
- I materiali consentono eventuali gradi diversi di fruizione da parte di categorie differenziate di utenza?
- La digitalizzazione renderebbe disponibile i materiali a una parte dell'utenza o della popolazione che altrimenti non li userebbe o che addirittura non utilizza affatto il patrimonio dell'Istituzione?

#### *Impatto sull'istituzione*

- L'istituzione possiede risorse sufficienti alla completa digitalizzazione dei materiali in

questione?

- I risultati del progetto hanno un'utilità immediata?
- Il progetto sarà di richiamo attirando nuova utenza o indirizzando quella tradizionale verso sezioni meno consultate del patrimonio dell'Istituzione?

### *Tutela*

- Gli oggetti che si intendono digitalizzare sono oggetto di consultazione frequente?
- La digitalizzazione fornirà un valido ausilio alla preservazione del materiale fornendo ad una gran parte dell'utenza un sostituto per la consultazione?  
Questo fattore ha due diverse e opposte ricadute sulla questione: da una parte potrebbe essere utile digitalizzare il materiale più utilizzato per preservarlo dall'eccessiva usura che deriva inevitabilmente da una frequente consultazione. Al contrario, la digitalizzazione può trattare materiale che l'Istituzione vuole promuovere in assoluto o nei confronti di una tipologia di utenza. La prima motivazione potrebbe riscontrarsi più di frequente in biblioteche che hanno una vocazione più "di ricerca" e che possiedono materiale "di pregio" o comunque non sostituibile utilizzato come fonte da numerosi utenti e per il quale una consultazione online potrebbe costituire una valida alternativa. La seconda motivazione, la digitalizzazione di materiale scarsamente utilizzato e per il quale si desidera incrementare l'uso o la conoscenza, risponde alle necessità di tutte quelle biblioteche che vogliono dare adeguata promozione ad un fondo specifico o ad una particolare tipologia di pubblicazioni presenti nelle proprie raccolte.
- L'oggetto presenta carattere di deperibilità o fragilità legata sia alla tipologia di supporto, sia al suo status?

## *Il diritto d'autore*

Il diritto d'autore consiste, secondo la legge 633 del 1941, nella protezione giuridica delle opere di ingegno, a carattere creativo, qualunque sia il modo o la forma di espressione.

Il diritto di autore si articola nei diritti all'utilizzazione economica e nei diritti morali.

Per l'accertamento del soggetto del diritto di autore è innanzitutto necessario definire chi è da considerarsi l'autore di un'opera. In primis chi viene presentato come tale (Art. 8. "È reputato autore dell'opera, salvo prova contraria, chi è in essa indicato come tale nelle forme d'uso").

Nel caso di opere di più autori tutti detengono i diritti (Art. 10. "Se l'opera è stata creata con il contributo indistinguibile ed inscindibile di più persone, il diritto di autore appartiene in comune a tutti i coautori").

Nel caso di miscellanee (Art. 7. "È considerato autore dell'opera collettiva chi organizza e dirige la creazione dell'opera stessa") o di edizioni critiche di opere fuori commercio si considera autore il curatore (Art. 85-quater. "2. Fermi restando i rapporti contrattuali con il titolare dei diritti di utilizzazione economica di cui al comma 1, spetta al curatore della edizione critica e scientifica il diritto alla indicazione del nome")

Nel caso in cui l'autore sia defunto, i diritti, nei 70 anni successivi passano agli eredi (Art. 115. "Dopo la morte dell'autore, il diritto di utilizzazione dell'opera, quando l'autore stesso non abbia altrimenti disposto, deve rimanere indiviso fra gli eredi per il periodo di tre anni dalla morte medesima, salvo che l'Autorità giudiziaria, sopra istanza di uno o più coeredi, consenta, per gravi ragioni, che la divisione si effettui senza indugio")

### *I diritti all'utilizzazione economica*

I diritti all'utilizzazione economica (include il diritto di riproduzione) di un'opera spettano all'autore o all'editore (nel caso in cui l'autore abbia ceduto in forma scritta tali diritti) per tutta la durata della vita dell'autore e fino ai settanta anni successivi alla sua morte (L. 633/41 art. 25 "i diritti di utilizzazione economica dell'opera durano tutta la vita dell'autore e sino al termine del settantesimo anno solare dopo la sua morte").

Spettano all'editore anche i diritti all'utilizzazione economica per le miscellanee (Art. 38. "Nell'opera collettiva, salvo patto in contrario, il diritto di utilizzazione economica spetta all'editore dell'opera stessa, senza pregiudizio del diritto derivante dall'applicazione dell'art. 7"), come anche per le edizioni critiche di opere fuori commercio; per le sole edizioni critiche di opere fuori commercio però la durata di tali diritti è limitata a venti anni (Art. 85-quater. "1. Senza pregiudizio dei diritti morali dell'autore, a colui il quale pubblica, in qualunque modo o con qualsiasi mezzo, edizioni critiche e scientifiche di opere di pubblico dominio spettano i diritti esclusivi di utilizzazione economica dell'opera, quale risulta dall'attività di revisione critica e scientifica. 3. La durata dei diritti esclusivi di cui al comma 1 è di venti anni a partire dalla prima lecita pubblicazione, in qualunque modo o con qualsiasi mezzo effettuata (1)")

I diritti di utilizzazione economica per le opere di

- Amministrazioni dello Stato
- Province
- Comuni
- Accademie
- Enti pubblici culturali
- Enti privati senza scopi di lucro



è di venti anni a partire dalla prima pubblicazione, qualunque sia la forma nella quale la pubblicazione è stata effettuata.

Per

- comunicazioni
  - memorie pubblicate dalle accademie e dagli altri enti pubblici culturali
- i diritti di utilizzazione economica sono ridotti a due anni, trascorsi i quali, l'autore riprende integralmente la libera disponibilità dei suoi scritti.

Presso il Ministero per i Beni e le Attività Culturali è attivo un Registro Pubblico Generale delle Opere Protette che può essere consultato per verificare le opere protette e la data a partire dalla quale contare gli anni di tutela per l'utilizzazione economica.

I bollettini sono disponibili dalla pagina <<http://www.librari.beniculturali.it/genera.jsp?id=116&l=it>>. Alcuni autori ed editori hanno affidato alla SIAE la tutela dei propri diritti. Sul sito della Società Italiana Autori ed Editori è possibile consultare la lista degli autori ed editori per i cui diritti bisogna interpellare la SIAE. In caso di dubbio (ma anche per verificare l'aggiornamento di tale lista) occorre consultare la sezione OLAF della SIAE

Per autori ed editori che non siano registrati alla SIAE è necessario contattare l'editore stesso

#### *Opere orfane e fuori commercio*

Con l'espressione opere orfane ci si riferisce a tutti quei casi in cui non sia possibile venire a conoscenza o entrare in contatto con la persona o l'istituzione che detiene i diritti all'utilizzazione economica.

Le opere fuori commercio sono quelle per le quali non sono ancora scaduti i tempi che ne tutelano i diritti ma che non sono più disponibili in commercio e per le quali l'editore non ha intenzione di produrre nuove edizioni o ristampe.

9

Indubbiamente digitalizzare solo opere ormai escluse dai diritti d'autore è più agevole e meno costoso per le istituzioni ma in questo modo resta esclusa quasi la totalità delle opere del 20. secolo, porzione particolarmente significativa di patrimonio, soprattutto per ambiti per i quali una minima forma di aggiornamento è determinante.

A livello europeo,<sup>2</sup> vi sono numerose raccomandazioni affinché la digitalizzazione sia rivolta anche ad opere ancora sottoposte a diritto d'autore, in particolare se si tratta di opere orfane o fuori commercio.

Per quello che riguarda le opere fuori commercio, si dovrebbe creare un sistema di accordo con gli editori che preveda forme di licenza semplificata per le opere fuori catalogo e non più distribuite; l'editore ovviamente dovrebbe essere nelle condizioni di ricavare una qualche forma di vantaggio, seppur indiretto, nel veder digitalizzata una proprio pubblicazione.

In tal senso inoltre, la Regione Lombardia adotterà tutte le iniziative volte ad entrare in contatto in modo ufficiale con gli editori, in vece delle singole istituzioni.

Per quello che riguarda le opere orfane, l'Istituzione interessata alla digitalizzazione di un'opera dovrebbe preventivamente eseguire una ricerca "diligente" per stabilire chi ne detiene i diritti di utilizzazione economica. Nel caso in cui sia impossibile determinare ciò, l'opera dovrebbe essere inserita in appositi inventari on line che conterranno la segnalazione delle opere che siano state sottoposte (o lo siano ancora) ad acquisizione e da parte di quale istituzione.

---

<sup>2</sup> Commissione per il patrimonio culturale europeo. *Risoluzione del Parlamento europeo del 27 settembre 2007 su "i2010: Biblioteche digitali"*. <http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P6-TA-2007-0416&language=IT&ring=A6-2007-0296>

In questo modo verrebbe garantito anche il diritto dell'autore dell'opera orfana, autore non rintracciabile da parte dell'istituzione interessata alla digitalizzazione. Nel caso in cui volesse rivendicare quanto gli spetta, avrebbe immediatamente la possibilità di contattare chi ha effettuato la scansione.

#### *I diritti morali*

Il diritto morale resta sempre in possesso dell'autore, anche nel caso in cui abbia ceduto i diritti di utilizzazione economica. In concreto comprende

- il diritto di rivendicare la paternità dell'opera anche per opere pubblicate come anonime (art. 20 e 21);
- il diritto di opporsi a qualsiasi modifica dell'opera e a qualsiasi atto che possa danneggiarla, anche solo moralmente (art. 20);
- il diritto di determinare il momento di pubblicazione ed eventualmente di stabilire dei limiti alla pubblicazione (art. 24);
- il diritto di ritiro dal commercio dell'opera per ragioni morali (artt. 142 e 143).

#### *Contributi diversi*

Per tutti i documenti sottoposti a digitalizzazione che siano costituiti da contributi di tipo diverso (testo e immagini, ad esempio), è necessario verificare quale sia l'autore di ciascun contributo e se siano decaduti i diritti di utilizzazione economica.

In alcuni casi si può ipotizzare di oscurare nella visualizzazione su web i contributi (ad esempio fotografie) per i quali non si sia potuto assolvere al pagamento dei diritti e che comunque non considerino determinanti per la piena comprensione dell'opera.

## *I file immagine*

### *Le immagini digitali*

Nei progetti di digitalizzazione vengono generalmente creati almeno due tipi di file immagine:

file master: il file che viene creato direttamente dall'attività di scansione dell'originale.

file derivati: file creati da successive modifiche del file master e salvati in formati differenti e a risoluzioni più basse.

I file risultanti dai progetti di digitalizzazione cofinanziati dalla Regione Lombardia devono essere aderenti ai requisiti che sono identificati come standard e che sono ampiamente utilizzati in analoghi progetti.

Per i file master tali requisiti sono:

File formato TIFF 6.0;

Risoluzione finale effettiva di 400 dpi (la dimensione dell'immagine è quella del documento originale a 400 dpi);

Profondità del colore 8 bit scala di grigio per le pagine in b/n;

Profondità del colore 24 bit colore RGB mode per le pagine con elementi a colori.

La qualità delle immagini digitali è inevitabilmente legata anche a

11

- la qualità degli originali che si desidera riprodurre (che deve essere analizzata durante la ricognizione dei materiali da digitalizzare necessaria per avviare un corretto bando di gara);
- la grammatura della carta dell'originale (evitare effetti di trasparenze e sovrapposizioni di testi o linee);
- il prodotto che si intende realizzare:
  - un file immagine destinato alla sola visualizzazione via web
  - un file immagine destinato al salvataggio per poterne apprezzare i particolari
  - un file immagine destinato alla stampa
  - un file immagine destinato alla conservazione per la tutela dell'originale analogico
  - un file immagine corredato di testo (attività quest'ultima che richiede la scansione ocr)
  - ecc.
- l'immagine che si intende produrre: in b/n, in scala di grigio o a colori;
- il formato dei file;
- la dimensione dei file.

Bisogna tenere presente che una scansione di buona qualità produce inevitabilmente file di ingenti dimensioni; rinunciare a risoluzioni elevate comporta la perdita della visione di dettagli che possono essere, a seconda del materiale, di interesse.

### *Risoluzione e qualità*

Le immagini digitali sono foto elettroniche costituite da una griglia di punti chiamati pixel. A ogni pixel viene assegnato un valore tonale (nero, bianco, scala di grigio o colore) rappresentato da un codice

binario costituito da 0 e 1, che costituisce i bit. I bit che rappresentano le immagini sono spesso compressi; i computer interpretano questi bit producendo l'immagine o su video o a stampa.

La risoluzione indica il livello di dettaglio. Maggiore è la quantità di punti o pixel per inch, maggiore è la risoluzione di un'immagine. Questo valore si misura in dpi, quanto l'unità di misura sono i punti (dpi sta per *dots per inch*) o in ppi, quando l'unità di misura sono i pixel (ppi sta per *pixel per inch*).

Si parla di “risoluzione effettiva” quando la dimensione dell'immagine digitale corrisponde a quello dell'originale scansionato secondo i dpi previsti.

### *Quale risoluzione adottare*

La risoluzione adatta ad un'immagine di alta qualità è quella grazie alla quale è possibile apprezzare i dettagli minimi degli originali che si intende scandire.

Per avere la certezza di utilizzare la risoluzione giusta per le finalità di un progetto, generalmente si effettuano procedure comparative effettuando scansioni a diverse risoluzioni di un documento campione scelto tra quelli che rispondono ai requisiti del progetto. La scelta ideale dovrebbe cadere sulla qualità di scansione che sia in grado di cogliere il dettaglio in questione e che sia al contempo compatibile con le modalità di visualizzazione e con la velocità di connessione mediamente disponibile agli utenti (nel caso in cui le immagini siano diffuse online).

La risoluzione minima accettabile quindi dipende dal materiale che si intende acquisire e dal risultato che si vuole raggiungere.

La risoluzione scelta condiziona fortemente

1. le funzionalità implementabili
  - lenti di ingrandimento
  - ricerca testuale
  - ecc.
2. la modalità di fruizione da parte dell'utente
  - da web (valutando anche le capacità medie di connessione disponibili)
  - da supporto fisso (risorse consultabili solo da postazioni riservate interne ad un'istituzione)
  - solo da stampe (in caso di digitalizzazioni finalizzate alla tutela e non alla diffusione in via elettronica)
  - ecc.

E' necessario primariamente distinguere tra i file master e i file derivati; il file master è destinato solo alla conservazione e alla produzione di file derivati; per questo motivo è importante sia di qualità molto alta. I file derivati invece, quelli che sono effettivamente divulgati tramite web, dovranno avere risoluzione e quindi qualità, inferiore. I file TIFF a 400 dpi (formato e risoluzione consigliati per i file master) sono infatti troppo pesanti per poter essere visualizzati con le connessioni comunemente a disposizione; inoltre la grande dimensione dell'immagine digitale comporterebbe una visione molto problematica sui monitor.

Per tali motivi le risoluzioni effettive ideali per la messa a disposizione su Internet di file immagini in formato jpg (JPEG2000) sono

per documenti moderni contenenti testo a stampa:

72 dpi (qualità bassa)

150 dpi (qualità media)

300 dpi (qualità alta)

per i documenti contenenti sia testo che immagini:

150 dpi (qualità bassa)

300 dpi (qualità media)

per il materiale cartografico:

300 dpi (qualità media)

per le stampe:

300 dpi (qualità media).

### *Dimensioni delle immagini digitali*

#### Dimensione in pixel

Le dimensioni in pixel di un'immagine digitale si ottengono moltiplicando la larghezza e l'altezza dell'immagine per il dpi con il quale viene acquisita.

Ad esempio, se un'immagine 10x12 cm viene scannerizzata a 400 dpi si ottiene un'immagine digitale delle dimensioni di 4000x4800 pixel (ossia 10x400 e 12x400).

Un foglio in formato A4 avrà la dimensione in pixel di 8400x11800.

### *Modalità di riproduzione digitale dei toni*

Nella scelta della modalità di scansione è determinante scegliere tra b/n, scala di grigio o colore; ognuna di queste opzioni corrisponde non solo al numero di colori tramite il quale verrà resa un'immagine, ma anche al numero di bit (chiamato *bit depth*) che avrà l'immagine stessa.

#### Modalità b/n

I pixel che costituiscono l'immagine sono costituiti ciascuno da un bit che può rappresentare due soli toni (generalmente il bianco e il nero). È possibile selezionare preventivamente quale soglia del colore

originale deve essere considerata “nero”. Questa modalità è adatta per testi o per disegni lineari senza ombre e chiaroscuri.

#### Modalità scala di grigio

La scala di grigio può essere espressa tramite una bit depth da 2 a 8-bit. La resa a 2-bit permette una resa di 4 colori (dal bianco al nero, passando per due tonalità di grigio); la resa a 8-bit, di gran lunga la più usata, permette un rendimento a 254 toni di grigio, oltre al bianco e al nero, per un totale di 256 valori di luminosità.. Anche in questo caso è possibile tarare la percezione delle sfumature. Questa modalità è adatta a fotografie in b/n, illustrazioni con sfumature e materiale per il quale il riconoscimento dei diversi tratti di scrittura può essere determinante.

#### Modalità a colori

La scansione a colori a 24-bit (ma esiste anche a 32 o a 48) permette una riproduzione a 16.8 milioni di colori creati attraverso la combinazione di rosso, verde e blu (RGB). Questa modalità è adatta per tutti quei casi in cui l'originale ha una significativa presenza di colori.

Ovviamente la scansione di materiali diversi da quelli cartacei potrebbe richiedere la definizione di parametri del tutto specifici.

Il metodo di colore RGB, pur essendo standard, dal momento che viene sottoposto alla visione tramite applicazioni e monitor specifici può anche non rappresentare in modo esatto la gamma di colori prevista.

#### *Dimensione in byte*

Si può avere un'idea della dimensione in byte che avrà un'immagine digitale moltiplicando le dimensioni in pixel dell'immagine per la *bit depth* e dividendo il risultato per 8 (dal momento che le dimensioni delle immagini sono espresse in byte, ciascuno costituito da 8 bit).

È necessario inoltre ricordare che

- 1 Kilobyte (KB) corrisponde a 1.024 bytes
- 1 Megabyte (MB) corrisponde a 1.024 KB
- 1 Gigabyte (GB) corrisponde a 1.024 MB
- 1 Terabyte (TB) corrisponde a 1.024 GB

#### *Scelta del formato dei file*

Generalmente le immagini digitali si salvano nel formato TIFF (Tagged Image File Format) che, non essendo compresso, garantisce che l'informazione rimanga inalterata; da questa immagine master se ne derivano altre, spesso in formato JPEG (oggi anche JPEG2000), GIF, PNG, PDF, per renderle accessibili su web, oltre ad una miniatura (thumbnail) in GIF.

Nella scelta del formato da utilizzare dovrà anche essere tenuta in considerazione l'eventuale compressione; ad esempio i file JPEG hanno una compressione che comporta la perdita di alcuni dati.

Estensione	Significato	Descrizione	Pro/Contro
.tiff, .tif	TIFF (Tagged Image File Format)	File non compresso. Da 1 a 64 bit. Usato per le immagini ad alta qualità e per la conservazione delle immagini.	File di alta qualità ma che produce immagini molto pesanti e poco adatte alla diffusione su web. Molti programmi non riescono a rendere risoluzioni superiori a 24 bit.
.gif	GIF (Graphics Interchange Format)	File a 8 bit compresso. Supporta l'effetto di trasparenza.	File molto utilizzato per la diffusione su web. Il W3c sta cercando di sostituirlo con il .png.
.jpg, .jpeg	JPEG (Joint Photographic Experts Group)	File compresso. Da 8 a 24 bit.	Formato molto utilizzato su web e per la diffusione.
MrSid	Multiresolution Seamless Image Database	File compresso. Utilizzato dalla LC.	Permette la creazione di un unico file che contiene immagini a diverse risoluzioni. Capacità di compressione più alta rispetto al jpg.
.pcd	ImagePac, PhotoCD	File compresso. 24 bit.	Usato in particolare per le immagini custodite e diffuse su supporto ottico.
.png	PNG (Portable Network Graphics)	File compresso che sta sostituendo il formato .gif. 24 bit.	Raccomandato dal W3c. Non ancora supportato da tutti i software.
.pdf	PDF (Portable Document Format)	File non compresso. 4-64 bit. Usato in particolare per la diffusione di immagini di documenti.	Visualizzabile solo con un apposito software.
.pct	PICT	File compresso usato in particolare dai computer con sistema operativo Macintosh. Fino a 32 bit.	Supportato solo da alcuni software.

## Calibrazione

La calibrazione è un momento determinante nella definizione di un progetto. Per definire infatti dei parametri uniformi è necessario che l'ambiente nel quale viene compiuta la digitalizzazione, il monitor, le periferiche di acquisizione e le fonti luminose siano misurate in modo che possano essere valutate eventuali difformità che potrebbero occorrere durante la realizzazione del progetto.

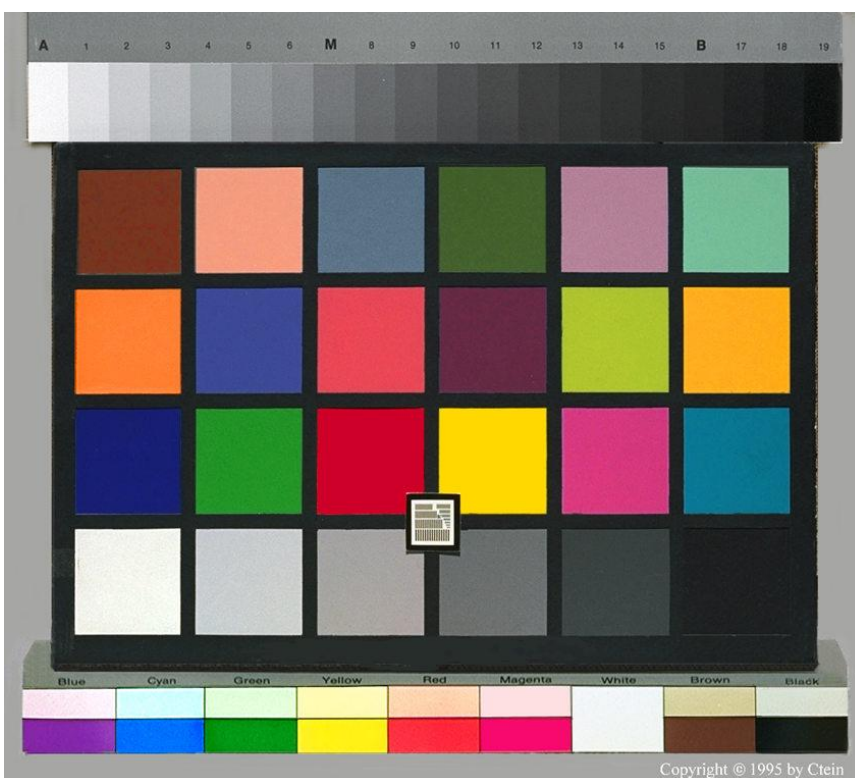
L'ambiente nel quale viene operata la digitalizzazione dovrebbe essere totalmente buio in modo che la sola fonte luminosa sia quella utilizzata dalla fotocamera; inoltre si dovrebbe evitare l'accumulo di polvere (soprattutto derivante dal materiale da scansionare!).

## Gestione del colore

Nel caso in si desideri una fedele riproduzione del colore è necessario ci si doti di un Color Management System (CMS), un insieme di prodotti hardware e software che servono per garantire un'uniforme resa dei colori su video, stampanti, ecc.

Nelle diverse fasi di trasferimento dei file da una periferica ad un'altra l'aspetto dei colori può variare; i diversi applicativi software che gestiscono le immagini e i monitor tramite i quali queste vengono visualizzate possono infatti avere modalità di definizione dei colori differenti.

Una soluzione che permette a bibliotecari e a utenti di verificare la corrispondenza dei colori consiste nel posizionamento, in uno dei file immagine che costituiscono la digitalizzazione di un oggetto, di un color checker (vedi immagine). In questo modo, grazie ad un parametro di riferimento costante, l'utente può comprendere gli esatti colori e tonalità delle immagini.





### *Inquadratura della pagina*

Nell'acquisizione dell'immagine corrispondente al foglio del documento, è necessario comprendere nell'area di scansione l'intero bordo del documento comprendendo anche qualche millimetro esterno allo specchio della pagina in modo da rendere l'idea dell'eventuale tridimensionalità del documento (nel caso di volumi). In questo modo l'immagine comprenderà sia una piccola parte del taglio delle pagine, sia della cucitura interna.

## Ocr e Codifica

### L'Ocr

Con l'acronimo Ocr, che sta per Optical Character Recognition, ossia riconoscimento ottico dei caratteri, ci si riferisce generalmente a quelle procedure e a quei software finalizzati alla conversione di un'immagine contenente testo in un testo digitale; in questo modo diventa quindi possibile passare da un file immagine (ottenuto tramite acquisizione digitale) a un documento modificabile e ricercabile tramite un normale editor testuale.

La qualità di un'acquisizione tramite ocr è influenzata da numerosi fattori: innanzitutto il carattere utilizzato nel documento, il contrasto del testo originale, la qualità del software di riconoscimento, il periodo di allenamento del software.

L'ocr presenta problemi particolarmente gravi per testi manoscritti e font che presentano caratteri fortemente ambigui per il riconoscimento.<sup>3</sup>

La Library of Congress ha stabilito che un percentuale di errore dello 0,5% è accettabile, a meno che non si stia producendo un documento di valenza scientifica; in questo caso infatti non è accettabile alcun errore. Lo 0,5% vuol dire una lettera errata ogni 200 (200 caratteri sono 2-3 righe di testo) Quindi, anche per i testi a stampa moderni è necessario che un operatore riveda il file prodotto dall'acquisizione, eventualmente prima attraverso il confronto con un correttore ortografico e successivamente tramite una verifica autoptica (un'altra possibilità potrebbe consistere nel far effettuare la scansione da software diversi, confrontando poi le versioni così ottenute).

### La codifica del testi

Con codifica o marcatura (da markup) si intende ogni mezzo per rendere esplicita l'interpretazione di un testo. L'introduzione di elementi esterni al testo (il markup appunto) permette di separare gli elementi logici del documento (codifica dichiarativa) e/o di specificare le elaborazioni a cui una determinata parte di testo deve essere sottoposta (codifica procedurale). Si tratta di istruzioni dirette alla macchina o all'uomo che permettono al testo di essere elaborato in più modi, ad esempio formattato secondo modelli differenti, indicizzato, estratto, selezionato, individuato.

La codifica procedurale, dunque, si occupa esclusivamente di come elaborare (processing - processare) il testo e di stabilire quale azione deve essere seguita in un particolare punto del documento. Al contrario la codifica dichiarativa non riguarda direttamente la formattazione testuale, ma la struttura logica del documento. In particolare si occupa di descrivere la struttura logica di un documento e l'informazione che contiene, dividendo il testo nelle sue parti costitutive, permettendo così all'elaboratore di comprendere tale struttura.

Scindere la struttura logica dagli aspetti meramente di formattazione testuale permette una maggiore flessibilità nella gestione del testo, ad esempio generando indici analitici automatici, aumentando le possibilità di ricerca verso tutti i segmenti strutturali riconosciuti dalla marcatura o permettendo di generare diverse visualizzazioni, diversi output, senza modificare la marcatura originaria.

Per fornire un esempio concreto, un documento marcato in TEI può generare (attraverso opportuni fogli di stile) un documento PDF, una o più pagine HTML etc. Inoltre allo stesso tempo può essere estratta la struttura del testo (capitoli – sottocapitoli etc) e possono essere generati indici analitici automatici. Infine la ricerca non sarà esclusivamente a testo libero (come nel caso di un semplice OCR), ma potrà essere diretta o limitata alle strutture logiche evidenziate dalla codifica.

---

<sup>3</sup> Per una recente presentazione del problema si veda M. Brantl – T. Garosci. *OCR: i progetti di digitalizzazione e il riconoscimento ottico dei caratteri*. Bollettino AIB 48/4(2008), 331-340. <http://www.aib.it/aib/boll/2008/0804331.htm>

## *Il controllo della qualità della digitalizzazione*

La qualità della digitalizzazione può essere verificata tramite il confronto con dati tecnici prefissati che devono essere compatibili con gli obiettivi che il progetto si pone di raggiungere. La qualità della digitalizzazione dipende anche dalla qualità dei metadati creati insieme al file digitale. Con l'espressione "controllo di qualità" ci si riferisce a tutte quelle procedure e pratiche utilizzate per assicurare la coerenza, l'integrità e l'affidabilità di un processo di digitalizzazione. Con l'espressione "garanzia di qualità" invece ci si riferisce generalmente alle procedure attraverso le quali si garantisce la qualità del prodotto finale.

### *File digitali*

Per quello che riguarda i file digitali la qualità può essere misurata su tre parametri:

- completezza  
il documento deve essere digitalizzato nella sua completezza, senza trascurare alcun elemento, incluse eventuali pagine bianche.
- leggibilità  
dalla visione del file digitale, che sia costituito da immagini o da testo digitato o catturato via ocr, l'utente deve poter comprendere il contenuto del documento originale. Questo comporta che l'immagine abbia un'adeguata risoluzione per la visione dei dettagli significativi del documento e che il testo sia comprensibile, anche nella sua struttura.
- fedeltà e rispetto dell'originale  
con fedeltà e rispetto all'originale si intende qualcosa che va oltre la completezza del documento: una corretta digitalizzazione dovrebbe permettere funzionalità avanzate, non realizzabili con l'originale, come la ricerca full text o l'ingrandimento di dettagli.

Se la completezza può essere facilmente misurata, la leggibilità e la fedeltà subiscono l'influsso di valutazioni sia di tipo soggettivo che oggettivo. Se le valutazioni soggettive, proprio a causa del loro carattere, possono variare da individuo a individuo, è necessario che ogni progetto definisca degli elementi di valutazione oggettiva come ad esempio, le procedure di analisi del testo ottenuto tramite l'ocr (su un campione del testo, ad esempio), il confronto tra la resa dei colori e l'originale (tramite ad esempio programmi di Color Management System).

Il controllo e la garanzia di qualità, elementi fondamentali di qualsiasi progetto di digitalizzazione, vengono spesso realizzati in modo informale, incoerente e casuale. È necessario invece che vengano specificati quali elementi delle immagini, dei metadati, dei testi, ecc. devono essere soggetti al controllo di qualità. Ad esempio per quello che riguarda la qualità delle immagini è facile stabilire parametri di confronto oggettivi.

Una volta stabiliti gli elementi e i parametri sui quali effettuare i confronti si dovrà anche definire se questa analisi andrà fatta in forma di campione o a tappeto su tutti i file digitali e in quale fase del processo di digitalizzazione; ovviamente tutte queste misurazioni andranno documentate.<sup>4</sup>

La verifica della qualità inoltre andrà ripetuta nel tempo; esiste infatti la possibilità che i supporti utilizzati per la conservazione sia dei file master che delle copie non abbiano doti di estrema longevità. Se gli hard disk tramite controlli periodici possono essere monitorati, i supporti come CD o DVD sono molto più fragili.

### *La qualità delle immagini*

---

<sup>4</sup> Precise indicazioni in tal senso sono presenti nel documento *Modello di capitolato*

Per ottenere delle immagini di qualità è necessario che la loro digitalizzazione avvenga su originali puliti, in un ambiente correttamente illuminato e tramite la strumentazione adatta.

Una serie di domande per il controllo di qualità delle immagini sono di facile definizione

- La pagina è stata catturata nella sua interezza o sono rimasti fuori o tagliati alcuni elementi come il titolo corrente o il numero della pagina?
- Sono state saltate alcune pagine dalla sequenza?
- È stato dato il nome corretto al file corrispondente alla posizione della pagina nella sequenza?

È necessario controllare, inoltre, per accertare la qualità del file digitale:

- la dimensione del file;
- la risoluzione;
- il formato di salvataggio;
- l'illuminazione;
- l'ombreggiatura;
- i dettagli dei chiaroscuri;
- il contrasto;
- eventuali pixel saltati;
- la centratura dell'immagine;
- la presenza di segni derivanti dalla scansione.

## *Metadati*

### *I metadati*

I metadati costituiscono un elemento fondamentale per i progetti di digitalizzazione; essi infatti assolvono a diversi scopi: descrivono la risorsa digitalizzata, ne illustrano la struttura e i diritti. È grazie ai metadati strutturali che viene definita la corretta sequenza dei file che costituiscono, nel loro insieme, un oggetto digitale. Sono invece i metadati tecnici che permettono agli utenti di comprendere i dati riguardanti il processo meccanico di digitalizzazione e, di conseguenza, verificare la rispondenza tra la versione originale e quella digitale, e a chi gestisce gli oggetti digitali, di effettuare le strategie per la conservazione.

### *Da cosa sono costituiti i metadati?*

I metadati appaiono come file di testo che includono, all'interno di etichette, i dati relativi a specifiche informazioni dell'oggetto digitale in questione.

Es. Intestazione (Header) di METS

```
<METS:metsHdr ID="BOOK1" CREATEDATE="2009-02-20T09:32:00"  
LASTMODDATE="2009-03-01T09:32:00" RECORDSTATUS="A">  
  <METS:agent>ROLE="CREATOR" TYPE="ORGANIZATION">  
    <METS:name>Biblioteca centrale della Regione Lombardia</METS:name>  
  </METS:agent>  
</METS:metsHdr>
```

21

---

### *Chi crea i metadati, e in che modo?*

I metadati possono essere creati in modo automatico, in modo semi automatico dal computer, oppure manualmente dal personale. In modo automatico possono, ad esempio, essere compilati campi come quelli relativi alla data di creazione del file o al suo nome; in modo semiautomatico, i dati tecnici relativi all'immagine o i dati bibliografici importati dal catalogo; in modo manuale invece il personale può descrivere la struttura del documento e la sua articolazione in capitoli o parti. Per creare quindi un documento METS si può usare un editor di testo, generico o specifico per XML, oppure, qualora si abbia a disposizione un apposito software, un'interfaccia che presenti un modulo da compilare e che quindi faciliti l'immissione dei dati.

### *Che rapporto esiste tra le immagini e i metadati corrispondenti?*

Le immagini e i relativi metadati sono archiviati separatamente, come distinti file, fin dal momento della loro creazione.

La scelta della Regione Lombardia per quello che riguarda lo schema di metadati da utilizzare è ricaduta su METS,<sup>5</sup> sviluppato nell'ambito della Digital Library Federation e mantenuto dal Network Development and MARC Standards Office della Library of Congress.

METS è

uno standard aperto  
non proprietario  
sviluppato dalla comunità bibliotecaria  
relativamente semplice  
estensibile  
modulare

METS consiste in uno schema scritto in XML destinato ad accogliere i dati descrittivi (ad es. titolo, creatore, anno, ecc.), amministrativi (ad es. diritti) e strutturali (ad es. per un libro digitalizzato, la corretta sequenza dei file immagine che corrisponda alla successione delle pagine nel documento cartaceo).

METS è solo una cornice, un framework, che non definisce le “regole” secondo le quali scegliere e immettere i dati.

Per questo motivo quando si parla di uso di METS si specifica anche quali standard si intende usare per “riempire” le diverse sezioni.

Un documento METS è costituito da sette sezioni principali :

1. Sezione Intestazione METS nella quale vi sono informazioni sul documento METS (l'istituzione o l'autore responsabile, la data di creazione del file, ecc.)
2. Sezione Metadati Descrittivi (*dmdSec*) nella quale è possibile o attivare un collegamento con una descrizione esterna (un record MARC, ad esempio), o inserire gli elementi descrittivi, oppure compiere entrambe le attività.
3. Sezione Metadati Amministrativi (*amdSec*) suddivisi in  
metadati tecnici (*techMD*) relativi alla compressione dei file immagine, alla profondità del colore, ecc.  
metadati relativi ai diritti (*rightsMD*)  
metadati relativi alla fonte (*sourceMD*), ossia all'oggetto digitalizzato  
metadati relativi alla provenienza digitale (*digiprovMD*).
4. Sezione File che presenta una lista di tutti i file che costituiscono l'oggetto digitale, anche riuniti in gruppi (*fileGrp*).
5. Sezione Strutturale nella quale viene delineata la struttura gerarchica che devono avere i file che costituiscono l'oggetto digitale per riproporre correttamente l'oggetto originario.
6. Sezione Link Strutturali, utile soprattutto nel caso in cui si trattino siti web.
7. Sezione Comportamento.

Grazie ai dati contenuti nelle diverse sezioni METS può essere utilizzato per

- la trasmissione e/o lo scambio di oggetti digitali tra repository
- l'archiviazione degli oggetti digitali per la conservazione e l'accesso
- la disseminazione degli oggetti digitali via web.

---

<sup>5</sup> Metadata Encoding & Transmission Standard <<http://www.loc.gov/standards/mets/>>

## *Strumenti tecnologici*

Una volta che siano stati definiti le finalità, i materiali da sottoporre ad acquisizione e i requisiti tecnici e descrittivi di un progetto di digitalizzazione, è necessario dotarsi della strumentazione tecnologica adatta a tali fini (o comunque, nel caso in cui la realizzazione del progetto sia affidata ad una ditta, verificare che ne sia dotata).

La scelta della periferica per l'acquisizione delle immagini dipende da diverse variabili: innanzitutto, se parte del progetto è affidato in outsourcing probabilmente la società alla quale verrà affidato l'incarico avrà già a disposizione alcuni prodotti. In questo caso si dovrà verificare se questi sono tecnicamente aggiornati o comunque compatibili con le finalità che si intendono perseguire e quindi con le specifiche delle immagini preventivamente stabilite.

Nel caso in cui la digitalizzazione avvenga invece all'interno della struttura si dovrà tenere presente dei seguenti fattori:

- Dimensioni massime che possono presentare i documenti che si intendono acquisire;
- Problemi relativi alla manipolazione degli oggetti;
- Competenze a riguardo del personale che lavorerà all'acquisizione delle immagini;
- Tempi di acquisizione previsti per ogni immagine (per verificare il cronogramma del progetto);
- Requisiti di qualità delle immagini;
- Budget.

### *La produzione dell'immagine digitale: la scansione*

La strumentazione di acquisizione deve garantire immagini che siano corrispondenti ai requisiti di qualità sovraelencati.

Il libro (i fogli sciolti presentano ovviamente meno inconvenienti) viene generalmente posto su un apposito supporto a forma di "v" che eviti sforzi nell'apertura; inoltre, in tal modo, la pagina mantiene la propria bidimensionalità senza deformazioni che ne altererebbero la corretta acquisizione.

L'acquisizione può essere comunemente effettuata tramite

una fotocamera digitale

uno scanner.

### *Le fotocamere digitali*

Le fotocamere digitali hanno il vantaggio di poter inquadrare secondo modalità differenti l'oggetto in questione, a seconda delle caratteristiche che questo presenta. Nell'uso professionale, la fotocamera non viene tenuta dall'operatore, ma è montata su una struttura che gli permette di assumere inclinazioni diverse. Nel caso in cui l'originale sia posto su un supporto a "v", spesso si fa uso di due fotocamere contemporaneamente, ciascuna orientata secondo le due angolazioni del supporto. In altri casi invece un'unica fotocamera viene spostata tra due posizioni o agisce grazie a un sistema di specchi.

La fotocamera inoltre permette all'operatore di verificare, sul visore della camera stessa o sullo schermo del computer, la correttezza dell'inquadratura e quindi la qualità dell'operazione.

La risoluzione di una fotocamera professionale deve essere di almeno 10-12 milioni di pixel.

## *Gli Scanner*

Riferendosi sempre a prodotti professionali si possono schematizzare prodotti e usi secondo la seguente tabella:

<b>Scanner planetari</b>	Documenti di pregio, manoscritti e stampati; documenti che presentano rilegature e/o condizioni di fragilità; grandi formati; quantità elevate di documenti.
<b>Scanner piani professionali</b>	Documenti cartacei di formato ridotto non rilegati; fotografie a stampa, pellicole e lastre fotografiche a media risoluzione.
<b>Scanner per diapositive</b>	Diapositive; pellicole 6x6, 6x9, 20x12 negative o positive.
<b>Scanner a tamburo</b>	Fotografie a stampa e pellicole fotografiche ad elevata risoluzione.
<b>Scanner a rullo</b>	Documenti cartacei a fogli singoli che consentono una acquisizione meccanica in quantità elevate.

### Gli scanner piani

Gli scanner piani (quelli comunemente presente in molte postazioni di lavoro) sono generalmente ritenuti inadeguati per i progetti di digitalizzazione in quanto:

- Il piano di scansione è di dimensioni ridotte o comunque fisse;
- Costringe un documento rilegato ad un'apertura eccessiva con possibili danni alla legatura;
- Comporta un'eccessiva manipolazione del documento;
- Acquisisce un'immagine deformata per l'effetto di curvatura delle pagine che si verifica vicino alla legatura;
- Non permette di verificare, se non a scansione avvenuta, la corretta posizione del documento.

### Gli scanner planetari

L'uso degli scanner planetari permette l'acquisizione di immagini fedeli agli originali, limitando in modo sensibile la manipolazione dei documenti. In alcuni casi viene posto un piano di cristallo a contatto con l'originale per livellare la superficie da acquisire.

Questa tipologia di scanner permette inoltre l'acquisizione di originali di dimensioni anche molto estese. Esistono anche scanner dotati di meccanismi meccanici che girano le pagine in modo automatico; tali strumenti sono molto utili in caso di campagne di acquisizione di ingenti dimensioni dal momento che permettono un notevole risparmio di tempo. Allo stesso tempo però possono essere usati solo per materiali di provata resistenza e che non presentino difformità nella dimensione dei fogli.

## *L'illuminazione*

La modalità di illuminazione incide sulla qualità della riproduzione inoltre può costituire un fattore di danno nei confronti del documento da acquisire. Gli illuminatori utilizzati nei progetti di digitalizzazione devono essere a luce fredda che, essendo priva di componenti IR (infrarossi) e UV (ultravioletti), non danneggia i documenti anche in caso di prolungata esposizione.



Anche la luminosità dell'ambiente dove si compie la scansione è importante per la qualità dei risultati.<sup>6</sup>

Alcuni scanner presentano il sistema di illuminazione incorporato nella strumentazione di acquisizione.

### *Altra strumentazione*

Oltre alle periferiche di acquisizione, sono ovviamente necessari computer e server che permettano la verifica da parte del personale dell'Istituzione delle immagini e dei relativi metadati (per il controllo di qualità) e la loro conservazione locale (a meno che questa funzione non sia delegata).

I computer in dotazione dovranno anche essere in grado di supportare il software per la gestione degli oggetti digitali.

Per la conservazione nel tempo è necessario avere a disposizione server di dimensioni adeguate allo spazio che serve per custodire i file immagine (master e derivati) derivanti dal progetto in corso e da altri eventualmente già in possesso dall'Istituzione.

---

<sup>6</sup> ISO 12646:2008 Graphic technology - Displays for colour proofing - Characteristics and viewing conditions  
<[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=44468](http://www.iso.org/iso/catalogue_detail.htm?csnumber=44468)>

## *Presentazione dei dati*

La presentazione dei dati è una questione di grande importanza; riguarda infatti le modalità tramite le quali un utente effettua una ricerca sugli oggetti digitali, visualizza i contenuti e naviga tra di essi, fattori determinanti per la comprensione degli oggetti digitali (e dei corrispondenti analogici) e per l'apprezzamento della loro qualità.

L'accesso agli oggetti digitali può essere

- veicolato ai record bibliografici presenti nel catalogo;
- reso disponibile attraverso specifici percorsi sul sito dell'Istituzione responsabile della digitalizzazione;
- reso disponibile attraverso la Biblioteca Digitale della Lombardia;
- reso disponibile dal portale di CulturaItalia.

L'utente potrà visualizzare e navigare gli oggetti digitali

- dal sito dell'Istituzione responsabile della digitalizzazione;
- dalla Biblioteca Digitale della Lombardia;
- ecc.

Le modalità di visualizzazione possono essere influenzate da una pluralità di fattori:

- il formato del file e l'eventuale forma di compressione;
- il browser utilizzato dall'utente per la navigazione;
- la capacità di connessione dell'utente;
- la qualità scelta dall'Istituzione per la diffusione delle immagini e le modalità con le quali l'immagine master è stata modificata;
- il computer (hardware e software) utilizzato dall'utente);
- ecc.

Ovviamente queste considerazioni danno in qualche modo per scontato che i fattori relativi alla qualità "logica" e "intellettuale" del progetto che ne garantiscono la coerenza e la logicità, siano già stati appianati.

*L'utente*

Al fine di progettare una corretta modalità di fruizione degli oggetti digitali, oltre alle numerose problematiche di tipo tecnico, è necessario tenere conto degli utenti che si presume potranno esservi interessati; per il Manuale Minerva,<sup>7</sup> utente è colui, professionista o no, specialista o no, che utilizza in modo casuale saltuario, finalizzato o sistematico l'applicazione web culturale pubblica.

In fase di progettazione è quindi necessario definire gli utenti per valutarne gli obiettivi e le possibili strategie e modalità di ricerca.

Il sistema dovrà comunque essere in grado di accogliere  
utenti automatici (motori di ricerca, harvester di metadati, ecc.)  
utenti interni (personale e amministratori)

---

<sup>7</sup> Minerva Working Group. *Manuale per l'interazione con gli utenti del Web culturale*.  
<[http://www.minervaeurope.org/publications/Handbookwebuserinteraction\\_it.pdf](http://www.minervaeurope.org/publications/Handbookwebuserinteraction_it.pdf)>

utenti simulati secondo profili (per luogo, età, istruzione, occupazione, lingua, uso di Internet, disposizione all'esplorazione ecc.).

Il Manuale di Minerva sulla User Interaction propone anche un questionario di autovalutazione<sup>8</sup> per coloro che intendono sviluppare un'applicazione web finalizzata alla ricerca e visualizzazione degli oggetti digitali.

---

<sup>8</sup> Minerva Working Group. *Manuale per l'interazione con gli utenti del Web culturale*. p. 148-152  
<[http://www.minervaeurope.org/publications/Handbookwebuserinteraction\\_it.pdf](http://www.minervaeurope.org/publications/Handbookwebuserinteraction_it.pdf)>

## *Conservazione digitale*

### *Cos'è la conservazione digitale*

La conservazione digitale consiste nel mantenimento delle capacità di visualizzare, recuperare e utilizzare le collezioni digitali a fronte dei cambiamenti tecnologici e organizzativi.

In particolare, comprende le seguenti attività:

- Mantenere aggiornati i file digitali, i metadati, i software utilizzati, tramite back-up e aggiornamenti periodici;
- Assicurarsi che le modalità di navigazione e ricerca offerte agli utenti siano sempre compatibili con i browser e con le loro nuove funzionalità;
- Garantire nel tempo la tutela dei dati;
- Garantire nel tempo la gestione degli eventuali diritti legati alle immagini.

### *Elementi critici*

I fattori di natura tecnica maggiormente critici sono:

- L'obsolescenza o il deterioramento (casuale o per negligenza) dei supporti;
- L'obsolescenza dei formati nei quali vengono salvati i file;
- L'integrità dei file stessi;
- Il cambiamento dei sistemi operativi, dei software e dei protocolli di comunicazione;
- Il cambiamento degli strumenti di recupero e di navigazione usati dagli utenti.

28

---

I fattori appena elencati vanno ad incidere negativamente sulla possibilità di visualizzare e gestire gli oggetti digitali nei casi in cui vi siano:

- Politiche di conservazione a lungo termine insufficienti quantitativamente o qualitativamente;
- Scarsità di risorse umane o economiche destinate a tale finalità;
- Disinteresse da parte degli amministratori nei confronti della gestione del progetto, una volta che questo sia stato avviato o terminato;
- Inadeguatezza dei metadati;
- Cambiamento delle leggi relative alla tutela del diritto d'autore al quale non corrisponda un adeguamento delle politiche di gestione degli oggetti digitali.

### *Strategie di natura tecnologica*

Vi sono diverse strategie che possono essere adottate per la conservazione del digitale; la letteratura scientifica sull'argomento ammette infatti l'impossibilità di determinare quale tra le soluzioni proposte si rivelerà nel lungo periodo la più affidabile.

Le nuove tecnologie, infatti, si modificano molto velocemente e secondo modalità non prevedibili.

Le strategie attualmente contemplate prevedono:

- il refresh periodico degli oggetti digitali tramite migrazione su nuovi supporti;
- il refresh periodico degli oggetti digitali tramite migrazione verso nuovi formati dei file immagine;
- la moltiplicazione delle copie relative agli oggetti digitali;
- il monitoraggio permanente degli oggetti digitali.

Queste soluzioni non si escludono a vicenda; le prime due soluzioni infatti impongono campagne di riversamento dei dati su nuovi supporti o di salvataggio dei file in formati diversi che avvengano con cadenza periodica nel tempo.

La moltiplicazione delle copie permette, nel caso in cui una risulti danneggiata o siano state fatte modifiche dannose o poco previdenti agli oggetti digitali, di avere a disposizione comunque un'ennesima replica.

Il monitoraggio permanente è la modalità attraverso la quale è possibile rendersi conto se uno dei fattori critici di natura tecnica elencati sopra minaccia gli oggetti digitali.

Il monitoraggio è una soluzione molto ardua da praticare da parte dell'istituzione: si tratterebbe infatti di dedicare delle unità di personale alla consultazione casuale degli oggetti digitali (che potrebbero essere anche un numero molto ingente) per verificarne la correttezza.

Il monitoraggio più efficace si può ottenere tramite la messa a disposizione degli oggetti digitali agli utenti sul web o, qualora non fosse prevista, sulla rete interna dell'Istituzione; la navigazione degli utenti è infatti il modo più naturale e più efficace per la verifica degli oggetti digitali.

Ad esempio, nel momento in cui una nuova versione di un browser non permettesse di visualizzare in modo corretto le pagine che presentano i risultati degli progetti digitali, gli stessi navigatori sarebbero incentivati a comunicarlo al responsabile del sito. Allo stesso modo, nel caso in cui un formato di salvataggio degli oggetti digitali diventasse obsoleto e quindi non permettesse più una corretta visualizzazione, gli utenti interni lo potrebbero fare presente all'istituzione.

Le alternative a queste pratiche sono:

- la creazione di programmi di emulazione che permettano l'utilizzo di formati "antiquati" dei file sui nuovi computer;
- la preservazione tecnologica;
- l'archeologia digitale.

Queste soluzioni, che mirano alla conservazione dei formati o della strumentazione adatta per la consultazione e la visualizzazione, sono costosissime nel lungo periodo e, comunque, non possono considerarsi soluzioni definitive. Arriverà infatti un momento nel quale non saranno più sostenibili economicamente o realizzabili.

### *Strategie organizzative*

Le soluzioni di natura tecnica per permettere la conservazione degli oggetti digitali possono essere intraprese solo a condizione che siano precedute ed affiancate da strategie di natura organizzativa, le più rilevanti delle quali riguardano:

- la formazione permanente;
- i finanziamenti;
- i metadati per la conservazione.

La formazione permanente è indispensabile affinché l'Istituzione possa disporre di personale in grado almeno di monitorare nel tempo gli oggetti digitali e di attuare le strategie di conservazione previste. Dal momento che la tecnologia si evolve è doveroso che, secondo modalità da stabilire, vi siano occasioni di formazione permanente che garantiscano l'aggiornamento delle conoscenze.

Le strategie di conservazione richiedono risorse finanziarie disponibili nel lungo periodo che non sempre possono essere preventivate in modo preciso; per questo motivo, l'istituzione che desideri intraprendere un progetto di digitalizzazione dovrà tenere in conto che, oltre all'impatto economico iniziale, vi saranno altri costi, non del tutto prevedibili, nel corso degli anni.

Un aiuto nella pianificazioni di queste attività viene dai metadati di conservazione, informazioni collegate ai file immagini che aiutano nel mantenimento della contesto nel quale gli oggetti digitali mantengono loro comprensibilità, autenticità, stabilità, ecc.

*Responsabilità locali/Responsabilità della Regione*

Si dovranno definire quali competenze in merito alla conservazione dei progetti spettano alla Regione e quali alle singole strutture, anche valutando che la Regione terrà una copia dei file prodotti per metterla a disposizione tramite il Portale.

## *Gestione del progetto*

Un progetto di digitalizzazione, oltre alla definizione dei parametri e delle procedure per l'avviamento e del lavoro, deve anche prevedere le modalità per la gestione in corso d'opera e per il mantenimento dei risultati nel tempo, una volta terminata la realizzazione.

Anche nel caso in cui si decida di affidare lo svolgimento del processo totalmente in outsourcing, all'istituzione spettano i controlli sulla qualità dei dati prodotti dal fornitore, da svolgere secondo il calendario stabilito nel capitolato, e la verifica della conservazione degli stessi nel tempo, insieme alla loro messa a disposizione degli utenti sul web.

Il processo di digitalizzazione non comporta infatti solo la creazione di file che costituiscono un valore di carattere intellettuale, ma produce anche un insieme di risorse di natura economica che spetta all'istituzione amministrare e gestire.

È necessario che vi sia un efficace sistema di gestione del progetto di digitalizzazione per

- Creare un efficiente archivio che custodisca le risorse digitali e i loro metadati
- Implementare un'infrastruttura che assicuri la gestione e la conservazione dei dati, garantendone l'aggiornamento nel tempo
- Fornire strumenti di ricerca che permettano agli utenti di identificare, localizzare e recuperare gli oggetti digitali.

A tal fine è necessario

- Coordinare le varie fasi del progetto;
- Gestire il personale coinvolto;
- Creare un sistema che
  - gestisca le parti amministrative (copyright, autorizzazione, sicurezza, misurazione dei progressi, ecc.),
  - garantisca un'efficace sistema di ricerca e di recupero dei dati,
  - verifichi che non vengano creati progetti simili che creino potenziali duplicati.

Gli strumenti messi a disposizione nel presente documento insieme al software opensource prodotto dalla Regione Lombardia hanno come obiettivo quello di aiutare e guidare l'istituzione in alcune fasi determinanti per il corretto svolgimento dei progetti di digitalizzazione.

## *Il personale*

Vi sono competenze diverse che contribuiscono all'avvio e allo svolgimento di un progetto di digitalizzazione e che devono essere individuate; alcune di queste mansioni possono anche essere svolte dalla medesima persona:

- responsabile dell'Istituzione (al quale spetta l'approvazione);
- responsabile del progetto (che ne segue lo svolgimento);
- personale con competenze tecniche per funzioni operative;
- personale con competenze tecniche per funzioni di controllo;
- personale addetto a funzioni non tecniche (spostamento dei materiali, ecc.);
- personale con competenze specifiche in ambiti collaterali (restauratori, ecc.);
- ecc.

La realizzazione di qualsiasi tipo di progetto digitale impone che quanti ne siano coinvolti siano in possesso di adeguate competenze e che ne curino l'aggiornamento.

Le conoscenze da acquisire variano in funzione a quali parti del progetto vengono svolte all'interno della struttura e quali vengono affidate in outsourcing. Ad esempio, nel caso in cui l'acquisizione delle immagini avvenga nell'Istituzione, il personale al quale verrà affidata questa mansione dovrà essere in possesso delle specifiche competenze tecniche.

Anche per la creazione dei metadati o per la codifica della struttura di un documento, nonostante parte di queste procedure potrà avvenire in maniera assistita, si dovranno fornire adeguate conoscenze.

Nel caso in cui l'Istituzione decida di utilizzare il software opensource messa a disposizione della Regione Lombardia, si dovrà provvedere alla formazione in merito al suo utilizzo.

Anche nel caso in cui un progetto sia affidato in outsourcing è necessario identificare all'interno dell'istituzione una o più persone che seguano il progetto.

In particolare è compito dell'Istituzione:

- preparare un elenco del materiale da digitalizzare che ne presenti, anche in forma sintetica, le caratteristiche tecniche;
- valutare le eventuali proposte derivanti da ditte esterne per lo svolgimento dei progetti di digitalizzazione;
- monitorare a campione, secondo cadenze prestabilite, l'attività di digitalizzazione, sia per quello che riguarda i metadati che per quello che riguarda la rispondenza ai requisiti tecnici previsti;
- gestire, per la parte che le compete, la conservazione digitale.

L'Istituzione dovrà quindi, per assolvere ai compiti sopra elencati, poter disporre di personale che sia in grado di analizzare

il materiale da acquisire individuandone e segnalandone le caratteristiche che possano essere rilevanti o problematiche ai fini della digitalizzazione;

le proposte delle ditte non solo da un punto di vista economico ma comparandone le offerte tecniche;

gli oggetti digitali creati verificandone la rispondenza con i requisiti richiesti;

nel tempo, gli oggetti digitali e le interfacce di navigazione e di consultazione per verificarne la correttezza e l'aggiornamento tecnologico.



Anche per quello che riguarda la formazione la Regione Lombardia potrà avere un ruolo determinante nel promuovere attività, coordinando gli sforzi ai fini di una ottimizzazione economica delle risorse.

## Flusso decisionale

Viene qui di seguito proposto un modello di flusso decisionale relativo ai progetti di digitalizzazione. Nello specifico, viene proposto uno schema che fa riferimento allo svolgimento in outsourcing di tale processo. A tale riguardo si rammenta l'importanza della verifica a campione della qualità dei dati (immagini e metadati), secondo quanto previsto nel *Modello di capitolato*.

