

Digitization of Ancient Documents and post-processing of digitized documents

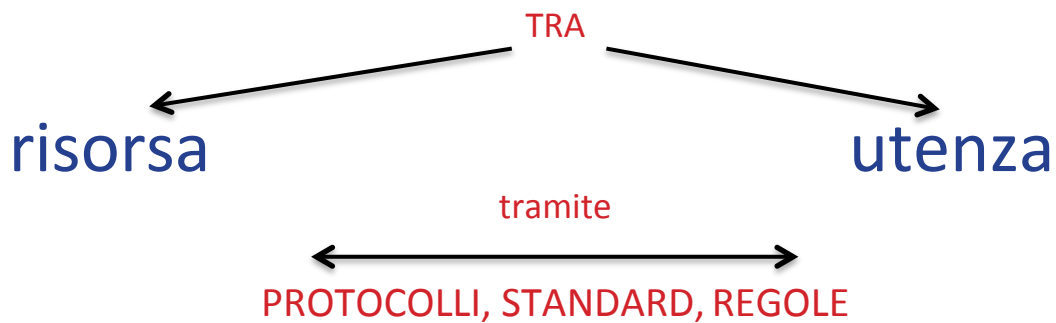
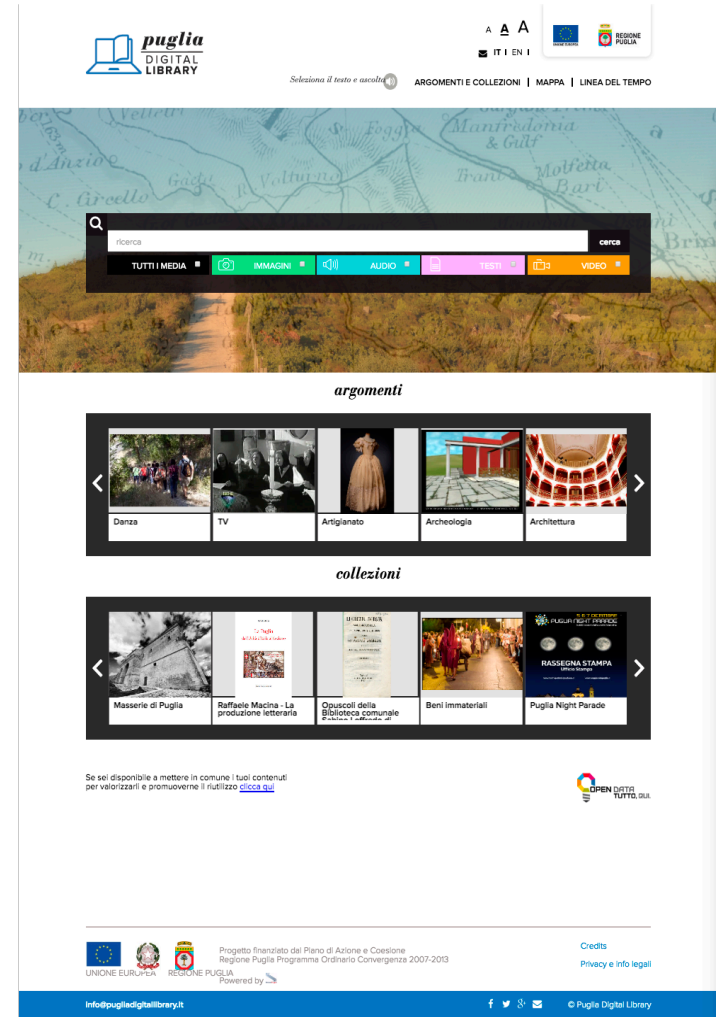
Digital Tools for Humanists Summer School
Pisa, 13 giugno 2019

Nicola Barbuti
Dipartimento di Studi Umanistici (DISUM)
Università degli Studi di Bari Aldo Moro

nicola.barbuti@uniba.it

BIBLIOTECA DIGITALE

infrastruttura per interazione

The screenshot shows the Puglia Digital Library website. At the top, there is a logo for 'puglia DIGITAL LIBRARY' and navigation links for 'ARGOMENTI E COLLEZIONI', 'MAPPA', and 'LINEA DEL TEMPO'. A search bar is visible with the text 'cerca'. Below the search bar, there are tabs for 'TUTTI I MEDIA', 'IMMAGINI', 'AUDIO', 'TESTI', and 'VIDEO'. The main content area is divided into two sections: 'argomenti' (topics) and 'collezioni' (collections). The 'argomenti' section includes categories like Danza, TV, Artigianato, Archeologie, and Architettura. The 'collezioni' section includes categories like Masserie di Puglia, Raffaele Macina - La produzione letteraria, Opuscoli della Biblioteca comunale, Beni Immateriali, and Puglia Night Parade. At the bottom, there are logos for the European Union and the Region of Puglia, along with a footer containing contact information and social media links.

BIBLIOTECA DIGITALE

requisiti minimi d'esistenza

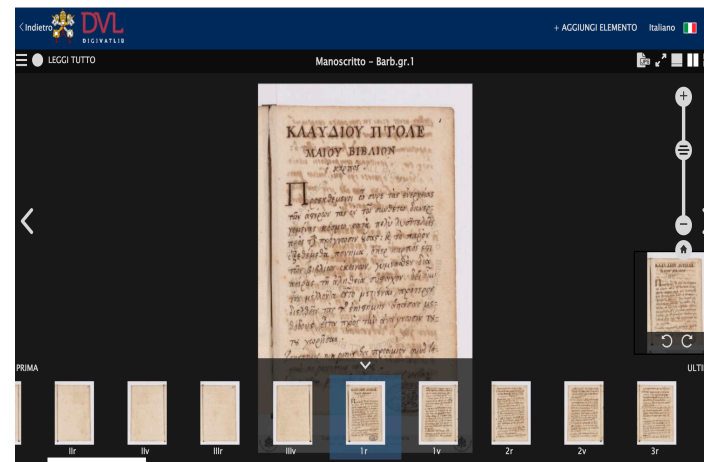
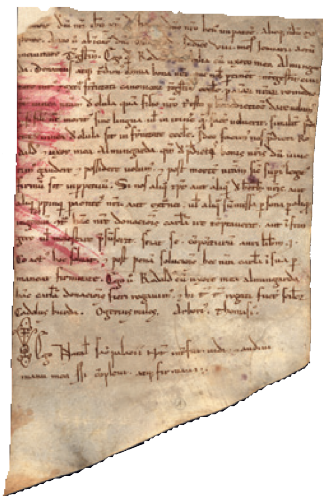
risorsa organizzata

PER LA MEMORIA

(conservazione)

PER LA FRUIZIONE

(pubblicazione/migrazione)



RISORSA ORGANIZZATA

requisiti minimi per la creazione



RISORSA ORGANIZZATA

significante dal processo di associazione tra

file

tif, jpg, pdf

metadati

DC, MAG, METS

STUDIO DI FATTIBILITA'

focalizzare il progetto

Comunità Montana del Lario Orientale

MODULO DI STUDIO BIBLIOTECOMICO E TECNICO DEL PROGETTO
"MONTAGNE DI SCATTI" (MODISCA) - CENTRO DOCUMENTAZIONE SULL'ALPINISMO
LECCHESE

Realizzato con il contributo finanziario della Regione Lombardia

Contenuti

0. Presentazione dello studio di fattibilità per il piano digitalizzazione di MODISCA

- 0.1. Introduzione
- 0.2. Struttura dello studio di fattibilità
- 0.3. Aspetti terminologici
- 0.4. Possesso e Accesso
- 0.5. Alcuni argomenti di partenza
- 0.6. La questione degli accessi
- 0.7. Il servizio di digitalizzazione online
- 0.8. Stabilire i parametri per i costi
- 0.9. Il Piano di sviluppo
- 0.10. Esempi di progetti di digitalizzazione presi in considerazione per il presente studio di fattibilità

1. Il piano di digitalizzazione di MODISCA

- 1.1. Scopo del piano di digitalizzazione
- 1.2.1. Criteri per la pianificazione
- 1.2.1.1. Linea 1. Principi per un accesso esteso all'informazione
- 1.2.2. Linea 2. Principi che tengano conto del modello di business per il servizio di digitalizzazione di MODISCA
- 1.2.3. Linea 3. I principi della cooperazione
- 1.3. Linee strategiche del piano
- 1.4. Obiettivi
- 1.5. Passaggio dal piano di digitalizzazione a programma permanente.
- 1.6. Lo staff e lo sviluppo delle raccolte digitalizzate
- 1.7. Documentazione e report
- 1.8. Amministrazione del programma
- 1.8.1. Il coordinatore
- 1.8.2. Standard in evoluzione
- 1.9. "Raccoglitori" di Buone Pratiche
- 1.10. Il ruolo dell'archivio
- 1.11. Negoziatori di contratti
- 1.12. Amministratore dei fondi
- 1.13. Promozione e marketing
- 1.14. Immagazzinamento
- 1.15. Azionario (compartecipazioni)
- 1.16. Ricerca e sviluppo
- 1.17. Strategie di migrazione (passare da/a).

- 1.18. Conversione a media futuri
- 1.19. Fondi per programmi di digitalizzazione
- 1.20. Entità singole/ collaborazioni
- 1.21. Sicurezza
- 1.22. Fondi esterni/ fondi locali
- 1.23. Risorse finanziarie

2. Il materiale da digitalizzare

- 2.1. Materiale della collezione digitale
- 2.2. La selezione delle collezioni
- 2.3. Individuare un nucleo di partenza dalle collezioni di MODISCA
- 2.4. Stabilire una massa critica per il modello di business

3. Approccio metodologico

- 3.1. Digitalizzazione: assetto preliminare al progetto
- 3.2. Coordinamento con gli altri progetti esterni di digitalizzazione
- 3.3. Livelli operativi dei progetti interni al piano: le sezioni
- 3.4. Linee e fasi di indirizzo
- 3.5. Le aree di intervento dello studio di fattibilità
- 3.6. Linee di intervento generate dalle tre aree
- 3.7. Gli strumenti del piano previsti nello studio di fattibilità

4. Creazione del registro digitale

- 4.1.1. La creazione del registro
- 4.2. Scopi del registro
- 4.3. Funzionalità degli registro
- 4.4.1. dati del registro: metadati per lo scambio
- 4.5. Dichiarazione dell'intenzione di digitalizzare
- 4.6. Informazioni provenienti da sorgenti multiple
- 4.7. Accesso
- 4.8. Visibilità come entità
- 4.9. Immissione dati e mantenimento

5. Indicizzazione e metadati

- 5.1. Indicare gli oggetti digitalizzati
- 5.2. Indicizzazione
- 5.3. Metadati
- 5.4. Indici e collezioni di metadati
- 5.5. Il processo di indicizzazione
- 5.6. Indice a modello centralizzato o decentralato
- 5.7. Ricerca dei metadati

- 5.8. Interoperabilità tra le raccolte
- 5.9. Elementi di metadati strutturali che sono richiesti per la conservazione delle copie master
- 5.10. Sistemi per la gestione di metadati

6. I criteri per i progetti entro il piano

- 6.1.1. Criteri per il progetto di digitalizzazione
- 6.2. La carta dei criteri per il progetto
- 6.3. Esempi pratici con dei costi

7. Pianificare le fasi di un progetto di digitalizzazione

- 7.1. La pianificazione
- 7.2. Selezionare e preparare i documenti da digitalizzare.
- 7.3. La creazione delle immagini digitali
- 7.4. Processo di codifica
- 7.5. Costi di digitalizzazione
- 7.6. Costi di post-digitalizzazione e altri costi aggiuntivi.

8. Linee guida per i progetti di digitalizzazione

- 8.1. Le cose da sapere
- 8.2. Scelte per l'accesso
- 8.3. Ottenimento dei diritti e dei permessi
- 8.4. Operatori specializzati/consulenti
- 8.5. Selezionare quali oggetti verranno digitalizzati
- 8.6. Acquisizione del materiale
- 8.7. Il processo di acquisizione
- 8.8. Componenti di un progetto di digitalizzazione
- 8.8.1. Progettare prima di digitalizzare
- 8.8.2. Natura della raccolta
- 8.8.2.1.1. documenti non elettronici.
- 8.8.2.2. Le immagini
- 8.8.2.3. Audio/video
- 8.8.2.4. Oggetti già in forma digitale
- 8.8.3.1. file digitali esistono in diverse forme.
- 8.8.3.1.1. Pointer records con immagini piccolissime.
- 8.8.3.2. Immagini di testo
- 8.8.4. Riconoscimento dei testi (OCR)
- 8.8.5. Copie per la conservazione
- 8.8.6. Media d'accesso
- 8.8.7. CD-ROM
- 8.8.8. Microfilm
- 8.8.9. Diapositive 35 mm
- 8.9. Ditte esterne rispetto a processi casalinghi
- 8.10. Catturare l'immagine
- 8.11. Scansione in loco rispetto a scansione centralizzata

STUDIO DI FATTIBILITA'

Valutazione della sostenibilità

Domande:

Quanto "peserà" il progetto?

Progetto chiuso, progetto aperto?

Quanti accessi?

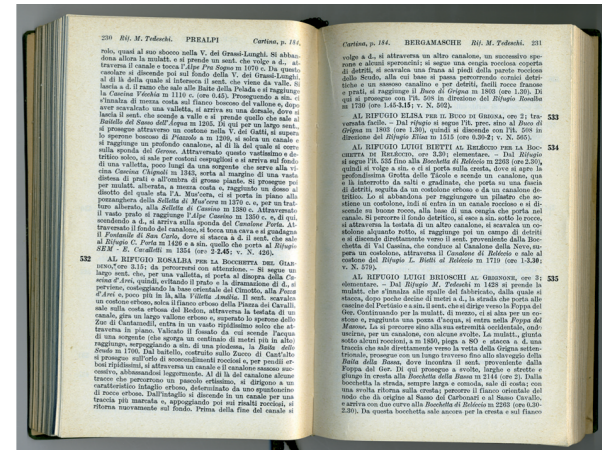
Quale capienza per lo storage?



Montagnedisconti

Centro studi
Archivio
documentazione
della "montagna" lecchese

Comunità Montana del Lario Orientale



Bibliografia di riferimento:

Gruppo di lavoro del Progetto Minerva, *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale*, Progetto Minerva, 2004.

http://www.minervaeurope.org/structure/workinggroups/goodpract/document/buonepratiche1_3.pdf

ANALISI DEI MATERIALI

- Caratteristiche fisiche
- Condizioni di conservazione
- Caratteristiche di tonalità e esposizione
- Presenza di dominanti di colore
- Nitidezza
- Tipo di supporto

Un preliminare accurato controllo qualitativo degli originali è fondamentale per l'ottimizzazione dei tempi di produttività.

ANALISI DEI MATERIALI

Strumenti analogici per l'analisi:

- Tavolo luminoso
- Piano illuminato
- Lentino o contafili
- Pompetta a bulbo, pennello a soffio e bomboletta di aria compressa
- Cartoncino grigio 18%

ANALISI DEI MATERIALI

La qualità della scansione dipende da quattro fattori:

- **Competenze dell'/degli operatore/i**
- Attrezzatura di acquisizione utilizzata
- Condizioni di conservazione e caratteristiche dell'originale
- Parametri ambientali

N.B.: nessuna operazione può compensare un originale di scarsa qualità. La valutazione è importante per definire il setting delle impostazioni dello scanner e gli eventuali interventi di post processing finalizzati a ridurre le problematiche degli oggetti digitali in produzione.

RISOLUZIONE DEI CONTENUTI

Le immagini digitali (*raster*) sono matrici di pixel.

Il pixel è l'unità più piccola attraverso la quale è possibile misurare l'informazione riprodotta in un oggetto digitale.

I pixel sono caratterizzati da quattro proprietà fondamentali:

- Dimensione (uguale per tutti i pixel contenuti in un oggetto digitale)
- Tonalità
- Profondità di colore
- Posizione

La dimensione è determinata dalla risoluzione cui l'oggetto digitale viene acquisito.

Es.: 600 ppi (points per inch) indica che ciascun pixel misura $1/600$ di pollice (1 pollice = 2,54 cm)

L'illusione dei toni continui si ottiene quando i pixel sono di piccola dimensione e i toni variano tra gli adiacenti secondo i parametri della scala cromatica cui appartengono i diversi insiemi.

RISOLUZIONE DEI CONTENUTI

La qualità della risoluzione è determinata dalla densità di informazioni digitali registrate in ciascun contenuto riprodotto.

Se impostata correttamente, contribuisce in modo significativo sulla percezione dei dettagli e ad assicurare che la tonalità sia fedele rispetto all'originale.



RISOLUZIONE DEI CONTENUTI

La risoluzione va definita in relazione alle caratteristiche fisiche dell'oggetto digitale e all'output che vogliamo ottenere.

Parametri cui attenersi per il setting:

- formato dell'originale o della porzione di esso che vogliamo acquisire.
- tipo e dimensioni dell'output finale
- profondità di colore

Requisiti tecnologici per la scelta e definizione della risoluzione:

- Ottimizzazione delle capacità dello scanner (evitare interpolazioni, scegliere risoluzioni integrali e fattore di ingrandimento integrale)
- Valutazione dell'incidenza della scelta della risoluzione sul workflow (manipolazione degli originali, dimensione dei file, post processing delle immagini, costo dell'output)
- Setting del monitor di visualizzazione

PROFONDITA' DEL COLORE

È il numero di bit utilizzati per rappresentare il colore di un singolo pixel.

Nel pixel la tonalità è indicata come combinazione dell'intensità dei tre colori fondamentali R(ed)G(reen)B(lue).

In base al numero di bit che viene assegnato a ciascun canale è possibile stabilire quanti valori diversi può assumere il colore.

Uno scanner in scala di grigi a 8 bit può acquisire 28 o 256 differenti valori di grigio.

Uno scanner a colori 24 bit acquisisce 28 livelli per ciascun canale RGB per un totale di $256 \times 256 \times 256 = 16.777.216$ colori

10.000/14.000 RGB x pixel: profondità del colore per pixel delle moderne CCD trilineari

Profilo ICC in uso: Adobe RGB 1998

STAZIONE DI SCANSIONE

Componenti hardware per la stazione di scansione:

- Scanner
- Workstation di gestione dello scanner (PC e monitor)
- Software gestionali
- HD esterni per storage e back up degli oggetti digitali

È fondamentale la qualità e l'efficacia tra questi elementi, al fine della massima funzionalità e produttività del sistema.

DISPOSITIVI DI ACQUISIZIONE

requisiti tecnologici

Il dispositivo per l'acquisizione ottica deve essere scelto in base ai seguenti fattori:

- Caratteristiche fisiche e stato di conservazione degli originali da acquisire (tipologie, condizioni fisiche, ecc.)
- Esigenze di output finale (pubblicazione sul web, stampa, video, applicazioni multimediali, OCR, archiviazione documenti, ecc.)
- Potenzialità di ogni tipo di dispositivo (produttività, area di analisi, ecc.)

Caratteristiche da verificare:

- Tecnologia dei sensori CCD
- Profondità di colore
- Risoluzione ottica (in ppi reali)
- Dinamica (intervallo di densità e massima densità)
- Formati originali accettati (per gli scanner)
- Formati d'immagine in uscita

DISPOSITIVI DI ACQUISIZIONE

tipologie

scanner

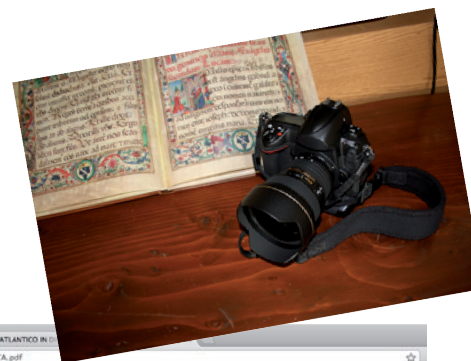
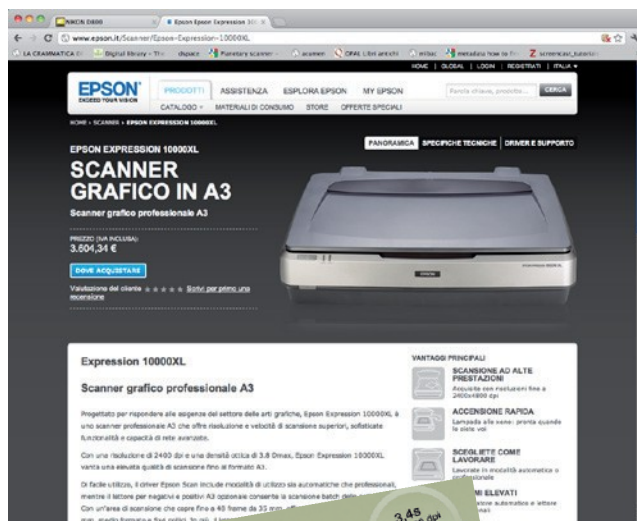
- a tamburo
- piani
- planetari
- per pellicole e diapositive

camere digitali

dorsi digitali

DISPOSITIVI DI ACQUISIZIONE

Scanner Planetari, piani, camere digitali



DISPOSITIVI DI ACQUISIZIONE

Planetari



Acquisiscono fino a ai formati A0

Per le loro caratteristiche tecnologiche sono particolarmente adatti alla scansione di artefatti documentali antichi.

Risoluzione: 300 – 1.600 ppi

Profondità di colore: 24 – 48 bit

Gamma di densità: 3,2–4

Motivi fondanti la scelta di scanner planetari:

- sicurezza nel trattamento degli originali
- elevata produttività
- eccellente qualità di acquisizione
- ottimo rapporto costo/beneficio



DISPOSITIVI DI ACQUISIZIONE

A tamburo



Alto livello qualitativo ed elevata produttività.

Utilizzano i PMT come lettori ottici.

Consigliabili solo per esigenze di grande precisione e di eccellente risoluzione, in quanto hanno costi eccessivi

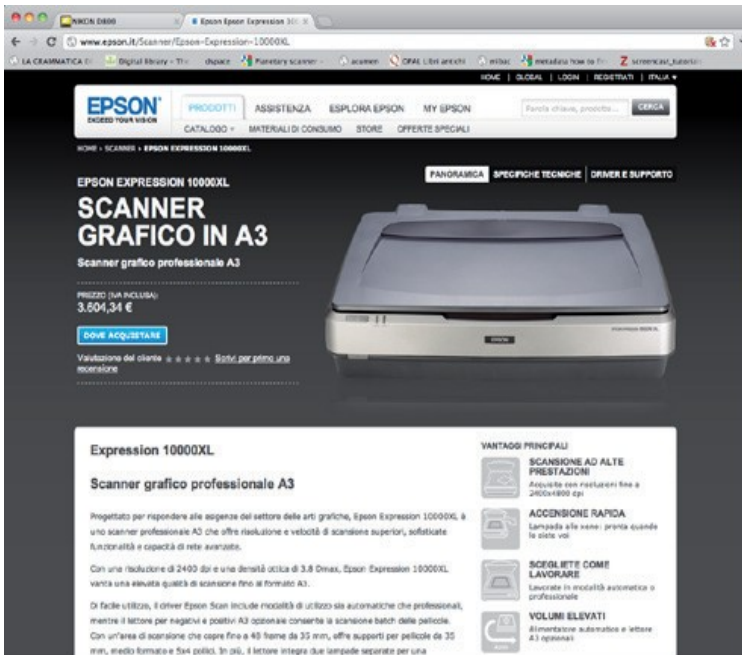
Risoluzione: 12.000

Profondità di colore: 30 – 48 bit

Gamma di densità: 3,6–4 (max 4,2)

DISPOSITIVI DI ACQUISIZIONE

Piani



EPSON EXPRESSION 10000XL
SCANNER GRAFICO IN A3
Scanner grafico professionale A3

PREZZO (IVA INCLUSA)
3.604,34 €

[DOVE ACQUISTARE](#)

Valutazione del cliente ★★★★☆ [Stati per primo una scansione](#)

Expression 10000XL
Scanner grafico professionale A3

Progettato per rispondere alle esigenze del settore delle arti grafiche, Epson Expression 10000XL è uno scanner professionale A3 che offre risoluzione e velocità di scansione superiori, sofisticate funzionalità e capacità di rete avanzate.

Con una risoluzione di 2400 dpi e una densità ottica di 3,8 Dmax, Epson Expression 10000XL vanta una silenziosa qualità di scansione fino al formato A3.

Di facile utilizzo, il driver Epson Scan include modalità di utilizzo sia automatiche che professionali, mentre il lettore per negativi e positivi A3 opzionale consente la scansione batch delle pellicole.

Con un'area di scansione che copre fino a 48 frame da 35 mm, offre supporti per pellicole da 35 mm, medio formato e 5x4 pollici. In più, il lettore integra due lampade separate per una

VANTAGGI PRINCIPALI

- SCANSIONE AD ALTE PRESTAZIONI**
Adatto con risoluzioni fino a 2400x1800 dpi
- ACCENSIONE RAPIDA**
Lampada alle xeni: pronta ovunque la si volesse
- SCEGLIETE COME LAVORARE**
Lavorate in modalità automatica o professionale
- VOLUMI ELEVATI**
Alimentatore automatico e lettore A3 opzionali

Grande facilità d'uso e costi accessibili.
I modelli recenti hanno prestazioni eccellenti, con risoluzioni ottiche molto alte.

Risoluzione: 300 – 5.000 ppi

Profondità di colore: 24 - 48 bit

Gamma di densità: 2,8–3,4

Assolutamente da non utilizzare per artefatti documentali antichi

DISPOSITIVI DI ACQUISIZIONE

Pellicole/Diapositive

Alta risoluzione ottica per un sufficiente ingrandimento di piccoli originali.

Supporti accettati: diapositive 35mm, pellicole 6x6, 6x9, 20x12 negative o positive.

Risoluzione: 1800 – 8.000 ppi

Profondità di colore: 24 – 48 bit

Gamma di densità: 2.5–4

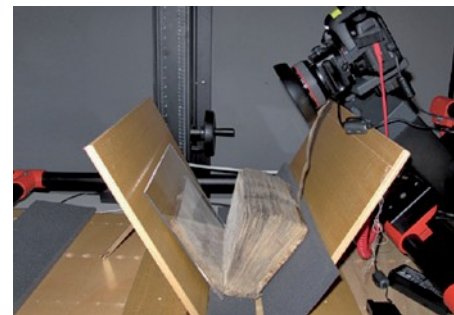
Costituiscono l'alternativa a basso costo per lo scanner a tamburo. Ottimo rendimento in combinazione con uno scanner piano.

Diapositive e le pellicole hanno un intervallo di densità di 2,8–3,0 quindi per acquisire tutte le tonalità esistenti tali scanner devono essere adottati di sensori CCD di alta qualità.



DISPOSITIVI DI ACQUISIZIONE

Camere digitali, stativo, set illuminazione



WORKSTATION

Elementi da valutare a completamento della stazione di acquisizione:

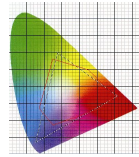
- Monitor e schede video (dimensioni, tipo, risoluzione).
- Potenza di calcolo (scheda madre, processore, RAM)
- Apparati per la memorizzazione di immagini e sistemi di trasferimento (HD, supporti rimovibili, dischi ottici, DAT, modem, ecc.)
- Strumenti per il quality control (color checker, etc.)
- Programmi (digitalizzazione, catalogazione, elaborazione delle immagini, calibrazione e gestione dei colori)

CALIBRAZIONE

Spazi di colore: mezzi attraverso i quali la scienza descrive il colore in termini concettuali e quantificabili.

Possiamo dividerli in:

- Spazi colore percettivi
- Spazi colore additivi
- Spazi colore sottrattivi



Obiettivo: rendere un'immagine finale identica all'originale.

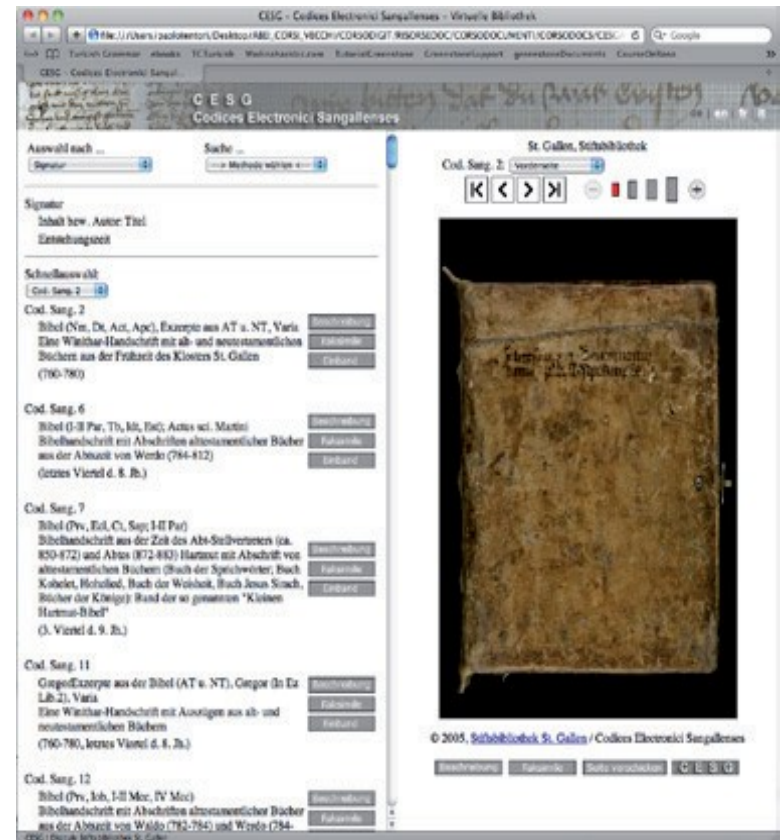
Possiamo individuare due fasi:

- Regolazione di ciascun dispositivo
- Armonizzazione delle caratteristiche di riproduzione del colore di tutti i dispositivi nel processo produttivo

I programmi specifici per la gestione del colore (CMS) consentono di accordare la riproduzione dei colori all'interno del flusso produttivo.

Processo di calibrazione:

- **Standardizzazione del colore dell'ambiente**
 - Luce costante
 - Colorazione neutra dell'area di lavoro
 - Immagini circondate da un colore neutro
- **Calibrazione del monitor**
 - Programmi o utilità di calibrazione
 - Calibratori basati su dispositivi hardware
 - Profili del monitor
- **Armonizzazione del colore scanner e altri dispositivi**
 - Calibrazione manuale (immagine campione)
 - Utilizzare un sistema di gestione del colore (CMS)



La gestione del colore avviene a livello di sistema operativo. Sia Windows che Macintosh supportano profili standard comuni proposti dall'ICC

PROCESSO DI ACQUISIZIONE

Fasi del processo:

- Predisporre la stazione di acquisizione: setting dei parametri di illuminazione, bilanciamento bianchi/neri, zoom ottico
- Selezionare la tipologia di originale le modalità di acquisizione, la destinazione
- Preparare e posizionare l'originale
- Effettuare un'acquisizione (pre-scansione) per il quality control dei parametri
- Se tutto coerente, eseguire la scansione

PRE-SCANSIONE

A bassa risoluzione, è preceduta un'anteprima da utilizzare per il quality control di scala di grigi e colori.

La pre-scansione viene utilizzata per le eventuali regolazioni di tonalità e il controllo dei margini.

Tagliare, centrare, ottimizzare sono operazioni da fare con attenzione e solo se necessarie.

Nessun taglio deve essere effettuato dall'operatore

I programmi di scansione di scanner di fascia alta presentano anteprime con più funzioni e a schermo pieno che facilitano il lavoro di regolazione.

FORMATI DEI FILE

Per la conservazione (in uscita dall'acquisizione):

- Tif 6.0
- FITS

Per la pubblicazione/fruizione:

- Tif 6.0 (copia per post processing)
- Jpg 300 (rete locale)
- Jpg 96-150 (web)
- Jpeg Thumbnails
- PDF

POST PROCESSING

Attività:

- Correzione del rumore
- Management del colore
- Correzione del colore e regolazione tonale
- Ritaglio
- Assemblaggio
- Riconoscimento digitale
- Optical Character Recognition (OCR)
- Intelligent Character Recognition (ICR)
- Intelligent Word Recognition (IWR)
- Pattern matching
- Graphic matching

Strutturare i metadati come risorse culturali

Metadata requirements for DCH identification, validation and certification: evolving R of FAIR Principles into FAIR⁴:

- Re-usable: reusability guarantees the sustainability of digital entities as the different re-uses over time determine their cultural memory function (an example above all: the Flavian Amphitheater, better known as the Colosseum);
- Relevant: relevance of a digital entity is characterized by the transformations of the function involved by re-use over time, and it is an indispensable requirement so that an entity, born with a purpose not necessarily cultural, evolves in memory and, therefore, in cultural resource by reusing several times with different functions;
- Reliable: reliability is connected to the digital entities representing within their content the validated and certified processes that have characterized its life cycle and evolution over time;
- Resilient: we can apply to metadata the resilience in its computational meaning of “the ability of a computer network to maintain service in the face of faults” (<https://en.wikipedia.org/wiki/Resilience>) turning it as “the ability of metadata to maintain its information over time and to guarantee both the knowledge of its original function and the capacity to provide a service even in the evolution of its functions from instrumental to cultural.”

E ripensare il processo di digitalizzazione

The R^4 requirements already shall be foreseen by the start of analysis and design of digitization process, whatever it is and of any complexity, whether it produces a simple digital object, or a digital library, or an immersive installation and so on, aiming at guarantee the capacity to evolve in memory.

The choice to give digital data the requirements of R^4 shall regulate both the methodological and technological approach to the process of digitization and to the co-creation of metadata structure and of each element description, with the goal to simplify the distinction among what can be digital cultural entities and can become Digital Cultural Heritage, and what just is consumer digital artifacts.

Creare digital cultural heritage

Our opinion is that the distinction between digital cultural entities and consumer digital artifacts lies in the quality of the co-design and drafting of descriptive metadata, and in particular in the correctness of the proportions between:

- quantity: it is the correct ratio between exhaustiveness of information, knowledge to be provided, number of metadata elements and attributes necessary to retrieve, reuse and store the digital resources, aimed at making them R^4 ;
- quality: it is the correct ratio between the informative/cognitive level to give both to each descriptor and to set of descriptors representing the resource and its life-cycle, and the variables of information and cognitive needs of the users, according to whether they are contemporary or future.

GRAZIE PER LA RESISTENZA