

Il presente documento ha come finalità la realizzazione di una Collezione digitale secondo gli standard nazionali e la sua fruizione dal portale italiano <http://www.internetculturale.it>

Le collezioni digitali possono essere affidate in gestione alla Teca Digitale Italiana (TDI) dell'ICCU, teca digitale per la fruizione da Internet Culturale.

I progetti destinati al sistema Internet Culturale devono essere realizzati secondo gli standard definiti dall'ICCU, che riguardano tutte le fasi del progetto, dalla scansione ottica dei documenti alla definizione e realizzazione dei metadati, obbligatori per la gestione e conservazione e necessari ai fini della ricerca sul web.

Per informazioni e documentazione si veda sul portale

e rivolgersi all'Istituto Centrale per il Catalogo Unico delle biblioteche italiane (ICCU) di Roma recapiti dei servizi digitali:

laura.ciancio@beniculturali.it ; marco.scarbaci@beniculturali.it

Vademecum per la digitalizzazione

Progetto e preparazione dei materiali:

Selezione dei documenti da digitalizzare finalizzata alla produzione di una collezione digitale, con caratteristiche tematiche, temporali, geografiche.

Conteggio dei file digitali che saranno il prodotto finale, finalizzato al calcolo dei costi di produzione, che vengono valutati, comprendendo anche i metadati, sul costo per ciascun file digitale. Si intende file digitale la singola riproduzione di pagina per la stampa, carta recto o verso per manoscritto, singola immagine, rimane doppia soltanto nel caso di necessità documentate.

Raccomandazioni per la produzione: file excel con elenco dei documenti da digitalizzare per collocazione/segnatura affiancata dal BID per SBN o CNMD per catalogo Manus. La cartella della digitalizzazione prende il nome del catalogo, come anche il file xml del metadato. Il riferimento del catalogo è sempre essenziale per la produzione digitale.

Vedere documenti pubblicati di “Specifiche file system” per i diversi materiali.

Scansione ottica

Le attività di scansione devono essere eseguite in sede in ambiente adatto con illuminazione costante e sempre uguale nelle diverse ore del giorno, con attrezzatura idonea, scanner planetario e server per la memoria di massa, nel pieno rispetto della normativa vigente in tema di sicurezza del lavoro, e per la sicurezza e conservazione dei documenti da trattare.

Tutta la strumentazione necessaria allo svolgimento del servizio viene installata e configurata dall'impresa, lo scanner deve essere calibrato sempre al montaggio, prima di procedere alle scansioni.

Tenuto conto del pregio e delicatezza degli originali, le modalità di ripresa che riguardano l'apertura dei volumi, l'illuminazione, la manipolazione, dovranno essere tali da non arrecare loro danno, le immagini saranno acquisite appoggiando i piatti della legatura sopra il piano basculante dello scanner, le pagine dovranno essere sfogliate con cura e senza esercitare pressione sul documento, in particolare sui dorsi delle legature. Durante tutte le operazioni di digitalizzazione il personale addetto dovrà indossare appositi guanti di cotone.

La riproduzione deve avvenire con l'utilizzo di lampade a luce fredda.

Verrà ripresa con la scansione unica la doppia immagine ma prodotto un file immagine (oggetto digitale) per ciascuna delle due parti relative a:

- ciascun verso e recto di ciascuna carta per i manoscritti; ciascuna pagina per testi a stampa, comprese le carte di guardia, anche se prive di informazioni, e le carte bianche sia interpolate che consecutive; i periodici vengono trattati a pagina e come fascicoli, non come volumi rilegati;
- tutte le parti componenti la legatura: piatti, dorso, tagli qualora significativi, (nel caso di codici, in modo da mostrare capitelli, fermagli, borchie, cantonali);
- la scala cromatica e scala millimetrica sarà posizionata una sola volta su una carta originale non significativa di ciascun documento; la carta da scansionare sarà concordata con il responsabile del progetto e posta come ultimo file immagine nel pacchetto digitale; la scala millimetrica deve essere posizionata lungo il bordo inferiore con lo "zero" allineato al bordo verticale della carta;
- le carte saranno riprese in modo "naturale", includendo i margini delle carte e delle parti del volume sottostante, per mostrare la materialità del documento nella sua interezza, senza interferire con la qualità dei colori sulla profondità di campo. Le riprese saranno effettuate su sfondo nero in modo da comprendere un'area minima scura, circostante la carta digitalizzata che ne consenta la visualizzazione nel contesto, comunque in modo tale che non venga mai esclusa dalla ripresa nessuna parte del documento. Le carte/pagine nella zona della legatura dovranno essere tagliate con un margine per mostrare la legatura e in piccola parte la pagina di fronte.
- in caso di presenza di lacerazioni, di fori di tarlo e ossidazione degli inchiostri, le carte devono essere mascherate con carta bianca in modo da evitare di riprendere il contenuto sottostante.
- In presenza di carte da riprendere, più piccole dei fogli sottostanti, porre al disotto del foglio in ripresa, una carta giapponese (non un comune foglio bianco), di spessore tale che consenta la visibilità delle pagine sottostanti in modo tale che non interferiscano con la lettura del foglio scansionato, di dimensione pari alle misure del documento.

La cartella delle scansioni sarà organizzata collocando nell'ordine: piatto anteriore, dorso, piatto posteriore, e in sequenza le carte interne, in fondo all'intero pacchetto la scala cromatica e millimetrica.

Formati di acquisizione immagini

L'acquisizione dovrà essere eseguita rispettando i seguenti parametri:

TIFF 6.0 non compresso ad alta risoluzione, 600 dpi ottici, profondità di colore di 24 bit RGB per formato inferiore o uguale ad A4; oppure 400 ppi, profondità di colore di 24 bit RGB per formato decisamente superiore ad A4.

Le acquisizioni dovranno essenzialmente garantire il rapporto 1:1 tra le dimensioni dell'oggetto analogico e quelle, "print size" (DPI), della versione TIFF dell'immagine

Tale digitalizzazione è destinata alla conservazione fuori linea e come copia di sicurezza, viene chiamata (master).

Dal master una volta verificata la qualità e completezza verranno successivamente prodotti:

JPEG in formato compresso a media risoluzione, 300 dpi ottici e profondità di colore di 24 bit RGB, destinata alla consultazione intranet (*non è obbligatorio solo se richiesto dal progetto*);

JPEG in formato compresso a bassa risoluzione, 150 dpi ottici, profondità di colore di 24 bit RGB, fattore di compressione 85 per cento, con ricampionamento (resampling) delle immagini, metodo tramite il quale il "print size" (DPI) del file JPG risulterà identico a quello del file TIFF.

Il valore del ridimensionamento della risoluzione dpi può essere aumentato a 200 dpi, o diminuito a 100 dpi, deve essere definito il suo valore in fase prototipale, in funzione di un'agevole consultabilità, a seconda della qualità del materiale (inchiostri, qualità della scrittura, della stampa, corpo dei caratteri), anche in relazione alle dimensioni dell'originale, sempre tenendo conto di non sottovalutare il troppo peso della cartella digitale finale ai fini del caricamento sul repository e della consultabilità da parte degli utenti. Particolare attenzione deve essere posta nella scelta dei dpi per i periodici, determinata dal corpo dei caratteri quando molto piccolo e dalle dimensioni del foglio; è infatti indispensabile una buona leggibilità a video con lo zoom.

Per i materiali di dimensioni molto piccole, di cui una delle due dimensioni risulta inferiore ai 10 cm (es. materiale fotografico), oltre a quello dei DPI, può essere opportuno applicare il parametro relativo alle dimensioni in pixel della risorsa digitale, definendo di default un valore per il lato lungo. Per Internet Culturale si suggerisce un valore di 2000 pixel per il lato lungo. In questo caso, non si dovrà applicare il metodo del ricampionamento, in quanto le dimensioni JPEG risulteranno necessariamente diverse rispetto a quelle della versione TIFF.

Tuttavia, una valutazione sull'utilizzo di questa modalità deve essere oggetto delle analisi, da parte di tutte le parti interessate, nella fase pre-prototipale.

Per materiali a stampa di qualità facilmente leggibile e per i periodici è necessario anche:

PDF con OCR embedded, tanti file quanti sono i file immagine, con la stessa nomenclature, sequence_number e nome del file del corrispettivo file IMG della copia master TIFF. La sez. OCR, avrà il solo usage 3. In sostituzione, è gradita la produzione di un unico file xml che elaborato in base allo schema ALTO di codifica, sia in grado di restituire le coordinate di ogni singola parola direttamente sul file di ciascuna immagine di usage 3 (sez. IMG per WEB).

Vedere per completezza della documentazione [Specifiche OCR](#)

E' sempre possibile il ricorso a programmi di miglioramento e fotoritocco per il formato JPEG per il web (correzione delle micro-rotazioni, rafforzamento del contrasto con filtri di smoothing e di riduzione del rumore, ecc.m soprattutto per digitalizzazioni in presenza di supporti particolari, come il lucido e la scrittura a lapis).

Prototipo delle scansioni e controlli di qualità

Il sistema di controllo della qualità delle immagini è finalizzato ad assicurare la buona leggibilità a video di tutto il contenuto informativo presente negli originali e ad evitare contestazioni a posteriori.

Il sistema che rende efficace la produzione è la creazione di un prototipo delle scansioni da produrre sulla base di un campione significativo dei materiali fornito dalla ditta. Significativo non nella quantità ma nella proposta di esempi per tipologia di materiali e/o con caratteristiche particolari (carte rovinare, scritte deboli, inchiostri acidi, caratteri piccoli o stampe deboli, ecc);

Il prototipo che individua i parametri/formati più opportuni, sulla base di quanto indicato nel precedente paragrafo, viene approvato con verbale dal responsabile del progetto per la Biblioteca e dal responsabile per l'Impresa, e costituisce il riferimento di qualità per le attività di scansione.

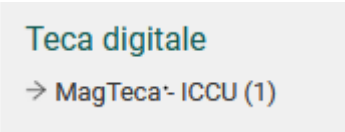
Identificazione delle immagini nelle memorie di massa

Nomi file delle immagini digitali

Durante la realizzazione del prototipo di cui sopra, il responsabile del progetto per la Biblioteca fornisce all'Impresa la lista dell'elenco dei materiali da digitalizzare con i codici BID/CNMD per la creazione delle cartelle immagini che avranno per nome il codice BID per SBN e il codice CNMD per Manusonline, e in alternativa è ammesso il codice Anagrafe della biblioteca_inventario_numero progressivo. Vale la pena di ricordare che non è ammesso il carattere blank.

Per i progetti digitali in consegna alla TDI dell'ICCU, verificare prima della digitalizzazione se presente altra copia con uguale identificativo nella TDI dell'ICCU. Come si verifica?

Scrivere il BID nel campo di ricerca semplice di Internet culturale,, con il puntatore sulla Biblioteca digitale. Se si ottiene un risultato di ricerca, scendere con il cursore e verificare nella faccetta Teca digitale, se la copia digitalizzata appartiene a MagTeca – ICCU.



Teca digitale
→ MagTeca- ICCU (1)

Solo in questo caso la regola della nomina dei file immagini cambia.

Se presente già una copia, aggiungere al BID_inventario, altrettanto dovrà essere fatto nel metadato, al nome dell'xml e al tag identifier.

Qualora si digitalizzino più copie del BID della stessa biblioteca, adottare la stessa prassi: BID_inventario.

Nelle memorie di massa, i file-immagine saranno distribuiti in più cartelle, almeno una per ciascuna unità catalografica (BID), anch'esse denominate secondo il procedimento appena esposto, in modo da preservare l'organizzazione complessiva dei materiali. Vedere [Specifiche file system manoscritti](#) ; [Specifiche file system monografie e periodici](#).

Nel caso dei Periodici, essendo complessa la casistica della periodicità che può essere usata per la preparazione delle cartelle, si invita a prendere contatto con gli uffici dell'ICCU, per entrare nello specifico dei singoli casi. In ogni modo i metadati dei periodici saranno realizzati per singolo fascicolo. Vedere [Specifiche file system monografie e periodici](#)

Nel caso di materiale documentario d'archivio, come corrispondenza, contratti, disegni, ecc., per la complessità si invita a prendere contatto con gli uffici dell'ICCU.

Produzione metadati secondo lo Schema MAG 2.01

I metadati di Internet culturale sono scritti in formato xml e in base allo standard SCHEMA MAG, versione 2.01 Vedere <http://www.internetculturale.it/it/1131/linee-guida-e-standard> lo Schema e altre informazioni.

L'xml metadato secondo lo Schema MAG contiene le sezioni

GEN, BIB, STRU, a seguire le digitali IMG, OCR, DOC, ecc.

La sezione BIB contiene il DC con 15 tag e alcune qualificazioni. La sezione si ottiene decodificando con algoritmi la mappatura dal Catalogo SBN-UNIMARC e dal Catalogo Manus-TEI, mappature pubblicate su Internet culturale

<http://www.internetculturale.it/it/1132/documentazione>

Qui sono presenti mapping da Unimarc verso MAG, per Libro antico, Libro moderno, Musica manoscritta, Cartografia, da TEI per il catalogo Manus.

Prototipo metadati e controlli di qualità

Il sistema di controllo della qualità dei metadati è finalizzato ad assicurare l'acquisizione in TDI e la ricercabilità dell'oggetto digitale attraverso i motori di ricerca.

Creazione di prototipi dei metadati sulla base di un campione significativo dei tipi di materiali digitalizzati nel progetto, predisposto dalla ditta e validato dall'Ufficio del digitale dell'ICCU;

Le verifiche in fase di acquisizione in TDI, potranno dimostrare inconsistenza e non qualità dei metadati, l'Ufficio che avrà facoltà di chiedere la riscrittura del mag difettoso.

Il prototipo che individua i parametri/formati più opportuni, sulla base di quanto indicato nei mapping di riferimento, viene approvato e costituisce il riferimento di qualità per la produzione.

Le indicazioni di nomenclatura della sez. IMG, indicazioni per l'OCR, le sez. GEN, e BIB possono essere concordate con l'ICCU, per il rispetto degli standard e delle norme redazionali di Internet culturale, sono pubblicate alla pagina <http://www.internetculturale.it/it/1132/documentazione>

Predisposizione e Consegna degli oggetti digitali

Al termine dell'attività digitale e metadati, il prodotto finale viene consegnato su HD Desk con interfaccia USB, completo delle cartelle degli oggetti digitali dei TIFF, JPG (300), JPG (150), in tre copie*, insieme un breve rapporto descrittivo del lavoro svolto, di tutte le scelte tecniche operate, nonché l'elenco dei contenuti degli HD Desk.

*i supporti saranno di marca diversa almeno due, la copia di uguale marca contenente cartelle digitali (con la sola versione web delle immagini) e metadati viene consegnata all'ICCU per l'acquisizione in TDI.