

PHAIDRA

Linee Guida sulla digitalizzazione di documenti bidimensionali

Versione 2011
Lorisa Andreoli
Marina Cimino

Revisione 2022
Lorisa Andreoli
Gianluca Drago



Luglio 2022



Quest'opera è distribuita con Licenza Creative Commons Attribuzione – Condividi allo stesso modo 4.0 Internazionale.



Indice

Indice	2
Premessa	3
1 Obiettivi	3
2 Selezione dei documenti	3
3 Conservazione	4
4 Digitalizzazione	4
4.1 Digitalizzazione <i>in-house</i> e in <i>outsourcing</i>	4
4.2 Scelta dell'attrezzatura	6
4.3 Acquisizione digitale	7
File master	7
Flusso di lavoro per acquisizione digitale da fotocamera	8
Flusso di lavoro per acquisizione digitale da scanner	9
File derivati	11
Testi da sottoporre a OCR (Optical Character Recognition)	13
Documento archiviato in Phaidra come libro	13
Documento archiviato in Phaidra come PDF ricercabile	13
Documento archiviato in Phaidra come immagine o come PDF semplice e processato da Internet Archive	13
4.4 Nomenclatura dei file	14
Libri	14
Riviste	15
Foto, manifesti, mappe (non rilegate), pergamene e altro materiale a fogli sciolti	16
Materiale di archivio	17
4.5 Controllo di qualità	18
4.6 Memorizzazione e conservazione dei dati	18
5 Archiviazione in Phaidra	19
6 Approfondimenti	19
6.1 Progettazione	19
6.2 Conservazione	20
6.3 Digitalizzazione	21
6.4 Metadati	22
6.5 Guide di Phaidra	22
7 Contatti	23
Allegato 1. Specifiche per i file XML dei testi sottoposti a OCR	24
Allegato 2. Scheda progetto di digitalizzazione	26

Premessa

Attraverso strumenti e strategie diversificate e nel rispetto della vigente legislazione sul diritto d'autore, il Sistema Bibliotecario di Ateneo (da ora in poi "Sistema") si propone di preservare e rendere accessibili in rete sia singoli documenti sia alcune significative collezioni digitalizzate relative a un ampio spettro di discipline, nel solco della tradizione scientifica e sperimentale che da sempre ha caratterizzato il nostro Ateneo.

Per valorizzare il proprio patrimonio documentale antico e di pregio e soddisfare l'esigenza degli utenti di un accesso rapido e facile ai contenuti informativi digitali, il Sistema si è dotato di Phaidra, una piattaforma di gestione degli oggetti digitali con funzioni di archiviazione a lungo termine in grado di soddisfare gli obiettivi specifici di conservazione e accesso alle collezioni digitali.

Questo documento, in coerenza con le buone pratiche e standard nazionali e internazionali per una riproduzione di qualità dei documenti, indica l'insieme delle procedure per la digitalizzazione di documenti di formato bidimensionale, con particolare riferimento all'archiviazione degli stessi in Phaidra. Il documento è disponibile anche in lingua inglese e viene periodicamente aggiornato. L'anno di prima pubblicazione è il 2011, successivamente è stato revisionato nel 2013, 2014, 2017, 2019.

1. Obiettivi

Definire gli obiettivi di un progetto di digitalizzazione consente di stabilire fin dall'inizio il quadro gestionale del progetto. In linea di principio i progetti di digitalizzazione devono essere coerenti con uno o più dei seguenti scopi generali:

- valorizzare il patrimonio documentale dell'Università di Padova e di altre istituzioni cooperanti
- favorire l'accesso a documenti rilevanti per valore storico, scientifico e culturale o con caratteristiche di unicità o rarità
- migliorare i servizi verso gli utenti, con la possibilità di consultare risorse fisicamente lontane, poco accessibili o poco conosciute raccogliendole in collezioni virtuali
- ridurre la consultazione di documenti originali in condizioni di vulnerabilità (documenti antichi e di pregio, in cattivo stato di conservazione, di difficile manipolazione, molto richiesti)
- assicurare che il materiale documentale sia disponibile alle future generazioni
- sviluppare attività di collaborazione con altre istituzioni

2. Selezione dei documenti

Nel selezionare i documenti da digitalizzare bisogna porre attenzione alle leggi sul diritto d'autore e sulla privacy. È necessario prendere in esame la titolarità dei diritti e la presenza di dati personali e sensibili in ragione della destinazione finale della digitalizzazione. Per risolvere possibili criticità

sarà necessario ricorrere alle soluzioni opportune (ad esempio, richiesta di autorizzazione a pubblicare), eventualmente rivolgendosi a un supporto legale.

I documenti vanno scelti sulla base degli obiettivi e dei criteri di selezione¹ definiti dal progetto. Generalmente si escludono opere già digitalizzate in altre raccolte o già accessibili al pubblico mediante la rete al fine di evitare le duplicazioni e contenere i costi.

In certi casi (ad es. materiale eterogeneo, non catalogato o inventariato) è raccomandato effettuare un censimento dei documenti per rilevarne la quantità, la tipologia, il formato, lo stato di conservazione, il valore patrimoniale e tutte le altre informazioni che si ritengono utili. A tale scopo, la *Scheda di progetto* (cfr. Allegato 2) può essere utilizzata come riferimento. Oltre che per la digitalizzazione, tali informazioni possono servire anche per attività di conservazione e catalogazione.

3. Conservazione

La cura e la conservazione degli originali sono attività integrate con la digitalizzazione. La valutazione dello stato di conservazione degli originali deve essere effettuata prima della digitalizzazione da personale esperto; eventuali modalità di restauro saranno definite da professionisti qualificati.

L'intervento di restauro dei documenti deve essere autorizzato dagli enti competenti in materia a seconda del contesto territoriale in cui si opera².

4. Digitalizzazione

La digitalizzazione è il processo di trasformazione/conversione di un oggetto analogico (libro, disegno, fotografia, pellicola o altro) in un formato digitale, interpretabile da un computer.

La natura e le dimensioni degli originali determinano la scelta del sistema di ripresa, del sistema d'illuminazione e delle modalità di trattamento (trasporto, apertura dei volumi, manipolazione).

La qualità delle immagini definita nel progetto determina i requisiti hardware e software del sistema di ripresa, i tempi di acquisizione ed elaborazione delle immagini, l'occupazione di spazio nei supporti di memorizzazione da gestire e conservare.

4.1 Digitalizzazione *in-house* e in *outsourcing*

La scelta della digitalizzazione all'interno dell'istituzione (*in-house*) o l'affidamento del servizio all'esterno (*outsourcing*) deve valutare vantaggi e svantaggi delle due modalità.

¹ Si veda *Selection for digitizing: a decision-making matrix* <https://www.clir.org/pubs/reports/hazen/pub74/matrix/>, in Dan Hazen, Jeffrey Horrell, Jan Merrill-Oldham, *Selecting Research Collections for Digitization – Full Report*, 1998 <https://www.clir.org/pubs/reports/hazen/pub74/>

² Per quanto riguarda il territorio del Veneto si faccia riferimento alla [Soprintendenza archivistica e bibliografica del Veneto e del Trentino-Alto Adige](#) e alla [Soprintendenza Archeologia, Belle Arti e Paesaggio per l'area metropolitana di Venezia e le province di Belluno, Padova e Treviso](#).

	Servizio all'interno (<i>in-house</i>)	Servizio all'esterno (<i>outsourcing</i>)
Vantaggi	<ul style="list-style-type: none"> • controllo diretto di tutto il processo • imparare facendo (<i>learning by doing</i>) • migliorare i requisiti in corso d'opera piuttosto che stabilirli in anticipo • provvedere a sicurezza, maneggio adeguato e accessibilità dei materiali 	<ul style="list-style-type: none"> • l'istituzione paga il prodotto, a un costo normalmente stabilito per immagine • contenimento del costo e rischio limitato • il fornitore può trattare grandi quantità di materiale • il fornitore assorbe i costi di specializzazione, formazione e obsolescenza tecnologica
Svantaggi	<ul style="list-style-type: none"> • l'istituzione paga le spese al posto dei prodotti, come costi di formazione, di obsolescenza tecnologica e dovuti ai tempi morti • investimento nell'acquisto e nel mantenimento dell'attrezzatura • richiesta di personale specializzato • costo per immagine non definito 	<ul style="list-style-type: none"> • l'istituzione elimina una fase del processo; non sviluppa una conoscenza approfondita sulla digitalizzazione • problemi di sicurezza, trasporto e maneggio degli esemplari
Raccomandazioni	<p>Il servizio all'interno è indicato se:</p> <ul style="list-style-type: none"> • la collezione non può essere spostata all'esterno dell'istituzione • il lavoro di digitalizzazione è relativamente facile • si può fare affidamento su personale specializzato e attrezzatura già presente 	<p>Il servizio all'esterno è indicato se:</p> <ul style="list-style-type: none"> • gli originali non consentono la digitalizzazione all'interno dell'istituzione • la programmazione è su grande quantità nel breve periodo • ci sono vincoli di spazio, infrastrutture e personale

La digitalizzazione in *outsourcing* può essere eseguita nei locali della biblioteca o presso la sede della ditta selezionata.

Il flusso delle attività di digitalizzazione *in-house* comprende:

- esame degli aspetti tecnico-logistici
- definizione dei parametri di digitalizzazione
- indagine di mercato o gara di appalto per l'acquisto dell'attrezzatura necessaria
- formazione del personale e degli operatori coinvolti
- predisposizione del set di digitalizzazione
- preparazione dei documenti
- realizzazione di un prototipo
- digitalizzazione
- controllo di qualità, con correzione di difetti ed errori
- ricollocazione dei documenti

Il flusso delle attività di digitalizzazione in *outsourcing* comprende:

- esame degli aspetti tecnico-logistici
- definizione dei parametri di digitalizzazione
- preparazione di un'indagine di mercato o di una gara di appalto
- eventuale predisposizione del set di digitalizzazione
- preparazione dei documenti
- formazione degli operatori coinvolti nel controllo di qualità
- realizzazione di un prototipo
- digitalizzazione
- controllo di qualità, con correzione di difetti ed errori
- ricollocazione dei documenti
- consegna del prodotto finito
- controllo di qualità finale

4.2 Scelta dell'attrezzatura

Il sistema di acquisizione (sorgente di luce, ottica, sensore, software di acquisizione e di calibrazione) deve garantire la qualità di immagine richiesta dal progetto e assicurare di non danneggiare i documenti originali.

In particolare il sistema di illuminazione deve essere a luce fredda, senza emissione di UV e IR.

Per i documenti antichi o di pregio dovrà essere garantito l'utilizzo di un supporto atto a non danneggiare il documento (rivolgendo verso l'alto la superficie da digitalizzare e con piano basculante o supporto a V).

Si riportano di seguito alcune indicazioni generali sui sistemi di ripresa³.

Scanner a letto piano: per documenti a fogli singoli, o documenti rilegati che possono essere aperti con facilità: materiale a stampa (es. volantini, manifesti, musica a stampa, brochure), manoscritti (es. lettere), mappe in buone condizioni, stampe (es. incisioni, acqueforti, litografie), disegni a penna e inchiostro senza aggiunta di acquarello o tempera (es. vignette), materiale fotografico (es. stampe alla gelatina in bianco e nero e a colori, stampe all'albumina).

Scanner per pellicole, negativi e diapositive: per pellicole, negativi e diapositive.

Scanner planetario oppure fotocamera digitale su stativo: per volumi rilegati (es. libri, album, musica a stampa, atlanti), documenti fragili, quadri a olio, la maggior parte delle opere d'arte su carta (es. acquarelli, disegni), materiale grafico e opere d'arte realizzate con sostanze sfaldabili e friabili (es. pastelli, carboncini, matita morbida), acquarelli con stesura spessa, a tempera o con vernici, mappe grandi o fragili, manoscritti (es. diari rilegati, documenti ripiegati), pergamene, materiale fotografico (es. stampe di grandi dimensioni; procedimenti fotografici storici, come dagherrotipi o ambrotipi), materiale tridimensionale (es. tessuti, sculture, oggetti).

³ Per approfondimenti sui sistemi di acquisizione vedi Ministero della Cultura, [Linee guida per la digitalizzazione del patrimonio culturale](#), 2022 (Versione in consultazione 2022-2023), paragrafo D.1.A.

Il Sistema Bibliotecario di Ateneo utilizza, assieme a sistemi di ripresa convenzionali – fotocamera, scanner a letto piano, scanner planetario –, anche uno scanner robotizzato per grandi formati, realizzato come prototipo hardware/software da uno *spin-off* dell'Università di Padova⁴.

4.3 Acquisizione digitale

Le seguenti specifiche sono da intendersi come indicazioni generali, da adattare caso per caso al fine di realizzare il miglior compromesso tra qualità e costo; infatti, se da un lato la produzione di immagini di elevata qualità implica costi elevati di acquisizione e di gestione (strumentazione di fascia alta, personale qualificato, tempi di acquisizione e di copia, spazio su disco), dall'altro i parametri di digitalizzazione adottati devono essere sufficienti a rendere fedelmente il documento con il livello di dettaglio necessario.

La dimensione delle immagini, la densità di campionamento, la profondità di colore devono essere valutate non solo in base alla dimensione del documento originale, ma anche in base alla tipologia di documento e alle risorse disponibili.

“È bene tenere presente che sono molteplici i fattori che influenzano la qualità delle immagini: fra questi, oltre alla densità di campionamento, annoveriamo l'accuratezza del colore, la dinamica del sensore e la sua rumorosità.

*Fissare una certa densità di campionamento è dunque concettualmente sbagliato perché, a seconda del sistema di ripresa che si usa, a parità di pixel per pollice, la qualità finale dell'acquisizione può essere molto diversa.”*⁵

Il risultato della digitalizzazione è la creazione di file destinati alla lunga conservazione, i file “master”, e di file frutto di elaborazioni successive, i file “derivati”, destinati alla fruizione da parte degli utenti, generalmente via disseminazione web.

Relativamente ai formati di file suggeriti di seguito, le indicazioni riportate riguardano esclusivamente la produzione di file immagine generati dalla digitalizzazione di documenti testuali, grafici o tridimensionali. Invece i formati di file suggeriti per altre tipologie di materiali, ad esempio pellicole e registrazioni sonore, sono analizzati nel documento [Formati dei file raccomandati per l'archiviazione a lungo termine e per la disseminazione web in Phaidra](#)⁶.

File master

Il **file master** (detto anche “master di conservazione” o “master di archiviazione”) è il file che rappresenta la miglior copia prodotta da una digitalizzazione, dove con migliore si intende quella che incontra gli obiettivi di un particolare progetto. Questi obiettivi possono variare a seconda della tipologia dei documenti.

I criteri da utilizzare nella creazione del file master devono garantire la riproduzione fedele del documento in vista della sua conservazione digitale a lungo termine o della necessità di una stampa di alta qualità, assicurando di non dover ripetere la digitalizzazione in futuro.

⁴ Per informazioni sull'attrezzatura utilizzata dal Sistema Bibliotecario di Ateneo dell'Università di Padova si veda la pagina web [Attrezzature per la digitalizzazione](#) (richiede autenticazione).

⁵ F. Lotti, M. Lunghi, G. Trumpy, [Digitalizzazione di beni artistici e documentari. Manuale di procedure per un laboratorio fotografico digitale](#), 2009. In particolare, si veda il capitolo 4 “Densità di campionamento”.

⁶ <https://phaidra.cab.unipd.it/static/formati-file-compresso.pdf>

I file master sono destinati all'archiviazione a lungo termine e vengono archiviati così come li restituisce lo strumento di acquisizione, senza subire alcuna rielaborazione.

Raccomandazioni per l'acquisizione dei documenti:

- Il documento deve essere ripreso nella sua interezza. Attorno al documento, si lascia un bordo di qualche millimetro che permetta di leggere il contorno del documento.
- Per i libri verrà prodotto un file immagine per ciascuna pagina: ciascun verso e recto di ciascuna carta, comprese le carte di guardia, anche se prive di informazioni, e le carte bianche; tutte le parti componenti la legatura: piatti, dorso, tagli, (in modo da mostrare capitelli, fermagli, borchie, cantonali). Per le mappe, per le fotografie e per i materiali di archivio, il verso andrà digitalizzato solo nel caso in cui presenti delle informazioni.
- Se l'originale è montato su un supporto che riporta informazioni (es. una fotografia montata su cartoncino con il marchio del fotografo), la digitalizzazione deve comprendere anche il supporto.
- Ogni documento deve essere affiancato da scala cromatica, scala di grigi e scala metrica, poste all'esterno del documento riprodotto e all'interno dell'inquadratura complessiva. Nel caso di un libro è sufficiente riprendere le scale una volta sola, non in ogni pagina.
- In presenza di lacerazioni, di fori di tarlo, di ossidazione degli inchiostri o di carta semitrasparente le carte devono essere mascherate con carta bianca in modo da evitare di riprendere il contenuto sottostante.

Flusso di lavoro per acquisizione digitale da fotocamera

L'acquisizione dell'immagine digitale a partire da un documento fisico risulterà più o meno dettagliata in ragione del tipo di fotocamera e della qualità dei sistemi ottici utilizzati.⁷

L'acquisizione digitale può essere condotta secondo due differenti modalità:

1. fissando la densità di campionamento che si desidera ottenere e utilizzandola per tutte le digitalizzazioni indipendentemente dalle dimensioni del documento fisico. Ciò si traduce nel mantenere sempre la stessa distanza tra il sensore della fotocamera e l'oggetto da riprendere con il risultato di ottenere immagini grandi nel caso di documenti fisici di grandi dimensioni e immagini piccole nel caso di documenti di piccole dimensioni
2. sfruttando tutta la matrice del sensore variando la distanza tra il sensore della fotocamera e l'oggetto da riprendere in base alla dimensione del documento fisico, con il risultato di ottenere immagini che indipendentemente dalla dimensione del documento fisico avranno tutte la stessa dimensione in termini di pixel acquisiti, ma differente densità di campionamento

Il Servizio Phaidra adotta la seconda modalità di acquisizione digitale.

Una fotocamera digitale può produrre file di formato differente.

⁷ Per tipologie e caratteristiche delle fotocamere vedi Ministero della Cultura, [Linee guida per la digitalizzazione del patrimonio culturale](#), 2022 (Versione in consultazione 2022-2023), paragrafo D.1.A.7. Per il workflow con la fotocamera vedi paragrafo D.1.B dello stesso documento.

Il formato *raw* – definito anche “negativo digitale”⁸ – è normalmente il miglior formato prodotto dalle fotocamere. Ogni casa produttrice ha il suo proprio tipo di formato *raw* proprietario, con proprie specifiche, che dà luogo a file con estensione caratteristica (ad esempio .nef per le Nikon, .cr2 o .cr3 per le Canon). Esiste un formato *raw* aperto, il formato DNG, ma normalmente non è supportato dalle fotocamere.

I master in formato *raw* prodotti dalle fotocamere vanno conservati così come escono dallo strumento di acquisizione, cioè senza ulteriori elaborazioni. Tuttavia, essendo in formato proprietario, non adeguato per la conservazione a lungo termine, vanno convertiti nel formato *raw* aperto DNG, non compresso. I file nel formato *raw* proprietario a questo punto possono essere cancellati. Le impostazioni in macchina fotografica al momento della ripresa dovranno garantire una adeguata qualità di colore (ad es. spazio colore ProPhoto RGB o Adobe RGB e profondità di colore a 16 o 8 bit per canale)⁹.

Flusso di lavoro per acquisizione digitale da scanner

La parola scanner viene usata per indicare tecnologie molto differenti tra loro. Spesso vengono indicati come scanner anche sistemi di ripresa che in realtà utilizzano una fotocamera montata su stativo, ma in questo documento con "scanner" facciamo riferimento allo scanner piano¹⁰.

Di solito il miglior formato prodotto dagli scanner è il formato TIFF.

Le impostazioni dello scanner possibilmente devono assicurare i parametri di qualità definiti nella tabella qui sotto per i diversi tipi di documento.

⁸ A differenza dei formati JPEG e TIFF che memorizzano immagini che sono già state elaborate dalla fotocamera, i file *raw* catturano dati non elaborati o elaborati minimamente direttamente dal sensore della fotocamera. Poiché sono analoghi ai negativi su pellicola nel flusso di lavoro di un fotografo, i formati *raw* vengono spesso definiti “negativi digitali” ([Adobe Digital Negative \(DNG\) Specification Version 1.4.0.0](#)).

⁹ Lo **spazio colore** (o metodo colore o *gamut*) è la gamma dei colori che possono essere visualizzati a monitor o stampanti. La gamma più ampia di colori è rappresentata dallo spettro visibile in natura, contenendo tutti i colori che possono essere percepiti dall'occhio umano, ma i dispositivi di acquisizione (fotocamere, scanner) possono avere più o meno limitazioni nel riprodurre l'intera gamma cromatica, e così pure i dispositivi di riproduzione (monitor, stampanti), quindi utilizzano spazi colore limitati. Gli spazi colore più utilizzati per l'acquisizione digitale, in ordine di qualità decrescente, sono ProPhoto RGB, Adobe RGB (1998), sRGB. Quest'ultimo è uno spazio molto limitato, ma comunque sufficiente per la visualizzazione a monitor (che in genere, a differenza delle stampanti, non è in grado di rappresentare uno spazio colore più ampio).
La **profondità di colore** rappresenta il numero di sfumature che è possibile ottenere per ciascuno dei tre colori primari (rosso, verde, blu). In ordine di qualità decrescente esistono 3 livelli di profondità: a 32, a 16, a 8 bit per canale. Quest'ultimo livello è sufficiente per la visualizzazione a monitor.

¹⁰ Per tipologie e caratteristiche degli scanner vedi Ministero della Cultura, [Linee guida per la digitalizzazione del patrimonio culturale](#), 2022 (Versione in consultazione 2022-2023), paragrafo D.1.A.1 e seguenti.

TIFF master

Tipo di documento	Formato file	Colore	Risoluzione ottica ¹¹
Materiale grafico (Fotografie, Stampe, Disegni, Dipinti, Manifesti, Mappe, Carte geografiche...)	TIFF 6.0, non compresso	Profilo colore "Adobe RGB (1998)" a 24 bit (8 bit per canale). Per documenti in cui si richiede una qualità superiore: Profilo colore "ProPhoto RGB" a 48 bit (16 per canale)	Fino al formato A4: 600 dpi. Superiore ad A4: 400 dpi. Per grandi e piccoli formati aggiustare la risoluzione in modo da ottenere il risultato ottimale
Libri, riviste e manoscritti, rari o di pregio (ad es. illustrati o dipinti) oppure con scarsa leggibilità (caratteri sbiaditi, basso contrasto, note a margine a matita, macchiati)	TIFF 6.0, non compresso	Profilo colore "Adobe RGB (1998)" a 24 bit (8 bit per canale). Per documenti in cui si richiede una qualità superiore: Profilo colore "ProPhoto RGB" a 48 bit (16 per canale)	Fino al formato A4: 600 dpi. Superiore ad A4: 400 dpi. Per grandi e piccoli formati aggiustare la risoluzione in modo da ottenere il risultato ottimale
Libri, riviste, manoscritti, dattiloscritti e ciclo-stilati, non rari, né di pregio, ben leggibili	TIFF 6.0, non compresso	Profilo colore "Adobe RGB (1998)" a 24 bit (8 bit per canale) o Scala di grigi a 16 bit	Fino al formato A4: 400 dpi. Superiore ad A4: 300 dpi. Per grandi e piccoli formati aggiustare la risoluzione in modo da ottenere il risultato ottimale
Negativi, Diapositive in Bianco e Nero	TIFF 6.0, non compresso	Scala di grigi a 16 bit	Da 35 mm a 10x12 cm: 800-2800 con una risoluzione basata su 4000 pixel lungo il lato lungo. Da 10x12 a 20x25 cm: 800-1200 con una risoluzione basata su 6000 pixel lungo il lato lungo. > 20x25 cm: 800 con una risoluzione basata su 8000 pixel

¹¹ La **risoluzione ottica** si misura in dpi, dove dpi sta per "dots per inch", una definizione che fa riferimento ai metodi della stampa. Il termine per descrivere la risoluzione delle immagini visualizzate sullo schermo di un computer invece è ppi, o "pixels per inch", anche se le due definizioni vengono spesso usate in modo intercambiabile. Sono le dimensioni in pixel dell'immagine finale a determinare la qualità complessiva, quindi è necessario regolare la misura dei dpi in modo da ottenere le dimensioni in pixel desiderate. Ad esempio, la scansione di una foto da 10 pollici di lato su uno scanner piano a 300 dpi produrrà un'immagine di 3.000 pixel di lato, mentre una foto da 5 pollici dovrebbe essere scansionata a 600 dpi per ottenere la stessa dimensione di 3.000 pixel. Si noti inoltre che mentre in uno scanner piano i dpi vengono preimpostati dall'operatore in fase di ripresa quando si utilizza una fotocamera digitale ciò non è possibile, perché è la dimensione fisica del sensore della fotocamera (in pixel/megapixel) a determinare le dimensioni finali dell'immagine e la sua risoluzione. L'attribuzione di una risoluzione definita espressa in dpi a un'immagine ricavata da fotocamera ha senso se si vuole fornire una qualità di stampa o una qualità di visualizzazione a monitor determinate.

			lungo il lato lungo.
Negativi, Diapositive a Colori	TIFF 6.0, non compresso	Profilo colore "Adobe RGB (1998)" a 24 bit (8 bit per canale). Per documenti in cui si richiede una qualità superiore: Profilo colore "ProPhoto RGB" a 48 bit (16 per canale)	Da 35 mm a 10x12 cm: 800-2800 con una risoluzione basata su 4000 pixel lungo il lato lungo. Da 10x12 a 20x25 cm: 800-1200 con una risoluzione basata su 6000 pixel lungo il lato lungo. > 20x25 cm: 800 con una risoluzione basata su 8000 pixel lungo il lato lungo.

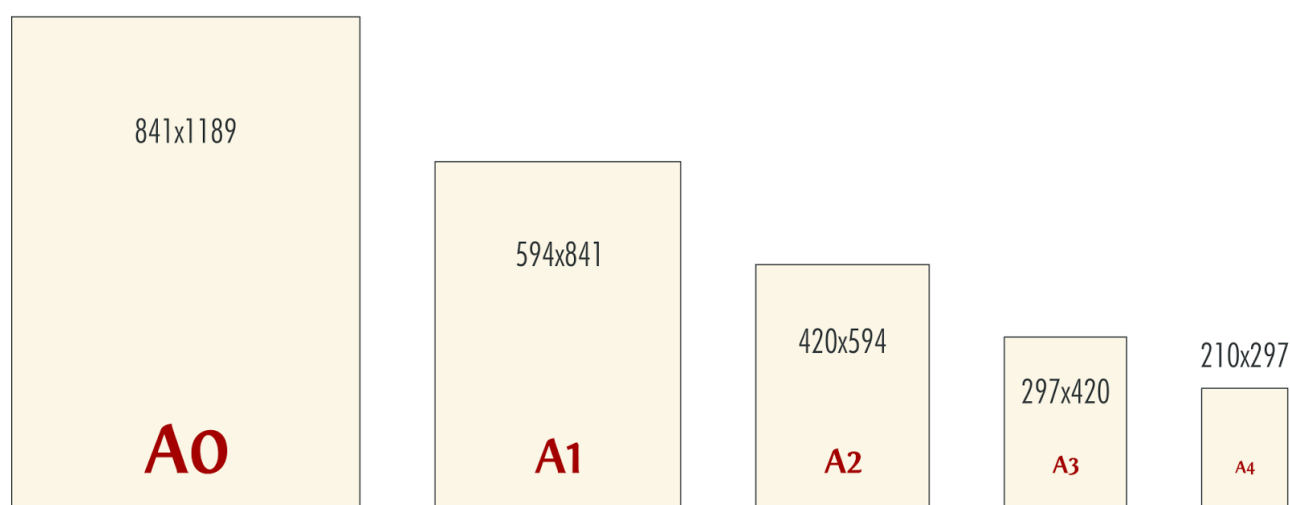


Illustrazione 1: Formato carta (mm). A0–A4 della Serie A, Standard internazionale ISO 216

File derivati

I **file derivati** sono prodotti a partire dal file master e ottimizzati per differenti fruizioni da parte dell'utente, ad esempio per la visualizzazione nel browser, per essere convertiti in testo via OCR oppure per la visualizzazione in una postazione dedicata. Sono normalmente ridimensionati e compressi, anche con perdita di informazione (ad es. immagini in formato JPEG), per la più comoda fruizione ottenibile senza un'eccessiva perdita di qualità.

La scelta della tipologia di file derivati da creare dipende dalle esigenze del progetto di digitalizzazione, tenendo conto della disponibilità di strumenti e di competenze che consentano di elaborare i file, dei differenti utilizzi previsti, nonché della qualità delle immagini che si desidera archiviare in Phaidra.

Di seguito vengono date delle indicazioni orientative sulle caratteristiche dei file derivati per i diversi casi di utilizzo.

TIFF. Nel caso di acquisizione da fotocamera in formato *raw*, dal file master si ricava un file derivato in formato TIFF 6.0 non compresso, con ordine dei byte Intel (PC); queste immagini dovranno essere fedeli all'originale e quindi non venire rielaborate in alcun modo, tranne che per il bilanciamento dei colori, operazione che si può fare con efficacia solo nei file *raw*.

JPEG. Sono file destinati alla disseminazione in web. Dovranno essere rimosse le scale cromatiche, di grigi e metriche presenti nel master. Inoltre, le immagini dovranno essere bilanciate per colore, luminosità, contrasto e saturazione al fine di correggere eventuali aberrazioni cromatiche dovute alle condizioni di acquisizione, sulla base dei campioni restituiti dalle scale di colore e di grigi. Il bilanciamento deve essere fatto a partire dai file master, prima dell'esportazione in JPEG, per via strumentale (con adeguati software di *color checking* e *photo editing*); deve puntare alla riproduzione fedele delle caratteristiche cromatiche dell'originale, non a un arbitrario miglioramento estetico. Le immagini dovranno anche essere raddrizzate e ritagliate per la visualizzazione ottimale.

Per quanto riguarda il ridimensionamento si dovrà operare una scelta di compromesso cercando di mantenere una buona qualità dell'immagine, ma senza produrre file eccessivamente pesanti.

In Phaidra anche immagini di dimensioni molto grandi (ad esempio con lato di 10.000 px) possono essere fruite agevolmente grazie al visualizzatore IIPImage¹² che consente di navigare e zoomare senza rallentamenti per l'utente. Tuttavia, bisogna considerare che il peso dei file influisce sia sui tempi di caricamento in Phaidra che sui tempi di scaricamento da parte dell'utente.

Per documenti che richiedono una visualizzazione di dettaglio molto elevata (ad esempio, carte geografiche oppure manoscritti antichi contenenti glosse minute) si consiglia di seguire le specifiche indicate nella tabella *JPEG di alta qualità*, mentre negli altri casi le immagini possono essere scalate a dimensioni inferiori rispetto a quelle dei master ad esempio seguendo le indicazioni presenti nella tabella *JPEG di qualità media*.

La risoluzione delle immagini è influente sulla visualizzazione a video, tuttavia se si desidera consentire una stampa di buona qualità si consiglia di utilizzare una risoluzione di 300 dpi.

La qualità di compressione JPEG deve essere valutata caso per caso, ma comunque non deve essere inferiore al 75%.

JPEG di alta qualità

Tipo di documento	Formato del file	Dimensione	Colore	Risoluzione ottica	Utilizzo
Tutti i documenti della tabella File master	JPEG, compressione di qualità medio-alta	Quella del master	Profilo colore Adobe RGB (1998) e profondità di 24 bit (8 bit per canale)	300 dpi	Mappe o altri documenti che richiedono una visualizzazione di dettaglio

¹² <https://iipimage.sourceforge.io/>

JPEG di qualità media

Tipo di documento	Formato del file	Dimensione	Colore	Risoluzione ottica	Utilizzo
Tutti i documenti della tabella File master	JPEG, compressione di qualità medio-alta	Riscalato a 3000 pixel sul lato lungo	Profilo colore Adobe RGB (1998) e profondità di 24 bit (8 bit per canale)	300 dpi	Documenti che non richiedono una visualizzazione di dettaglio

Testi da sottoporre a OCR (Optical Character Recognition)

Nel caso si vogliano mettere a disposizione dei file con testo ricercabile è necessario sottoporre le immagini digitalizzate a OCR.

In tal caso si può creare un PDF ricercabile, nonché altri formati di testo secondo le esigenze del progetto (TXT, EPUB...)

Documento archiviato in Phaidra come libro

Nel caso in cui si voglia archiviare in Phaidra un oggetto di tipo "Libro" rendendone ricercabile il testo:

- bisogna che l'OCR venga eseguito su immagini della stessa dimensione di quelle che saranno caricate in Phaidra
- per ogni immagine (quindi per ogni pagina del libro) bisogna creare un file XML, che abbia lo stesso nome del file immagine
- bisogna che l'XML segua la formattazione dell'esempio in *Allegato 1*
- bisogna creare un PDF ricercabile

Documento archiviato in Phaidra come PDF ricercabile

Un possibile flusso di lavoro per creare un PDF ricercabile viene descritto nel documento "Creazione di PDF ricercabili con software libero. Caso studio: i Bollettini-Notiziari della Facoltà di Scienze Statistiche"¹³.

Documento archiviato in Phaidra come immagine o come PDF semplice e processato da Internet Archive

Per gli oggetti che hanno una licenza che ne consente il riuso è stato predisposto un flusso di lavoro per esportarli in modo semi-automatico da Phaidra a Internet Archive ottenendo con questo un doppio risultato: una maggiore visibilità internazionale e la messa a disposizione in Phaidra dei file

¹³ Previa autenticazione: https://bibliotecadigitale.cab.unipd.it/collezioni_navigazione/Members/bibliotecari/materiali_settore_bd/gl-biblioteca-digitale/gruppo-phaidra/factory/creazionepdfricercabili.pdf

testuali creati da Internet Archive (PDF ricercabili, TXT...). Per poter utilizzare questa opzione occorre fare richiesta al Servizio Aiuto del Sistema bibliotecario¹⁴.

4.4 Nomenclatura dei file

In generale il nome di ciascun file sarà una stringa di caratteri composta da più parti, costruita per permettere di risalire alla collocazione del documento fisico dal quale deriva l'immagine. I nomi dei file saranno completati con l'estensione adeguata (.tif, .jpg, .pdf, .xml).

Nelle memorie di massa i file immagine saranno distribuiti in più cartelle, in modo da mantenere l'ordinamento complessivo dei materiali.

La nomenclatura delle cartelle e dei file è una stringa di campi (Codice Biblioteca¹⁵, Segnatura...) separati da trattino (-). Laddove la segnatura contenga già al suo interno dei trattini (-), o degli spazi o dei caratteri non alfanumerici, questi vengono rimpiazzati da un punto (.

Se utile, per materiale grafico e materiale di archivio che viene digitalizzato su entrambi i lati, si può far seguire alla numerazione progressiva dei file “-r” per il recto, “-v” per il verso.

Per i libri, coperta anteriore e posteriore vengono nominate in modo che si presentino nello stesso ordine che hanno nel documento fisico. Dorso o altre parti del documento originale (tagli, dettagli della legatura...) devono presentarsi alla fine.

Nelle immagini che includono le scale di colore e di grigio e il righello alla numerazione progressiva del file si fa seguire “-c”.

Libri

La cartella principale, denominata “*Codice Biblioteca – Segnatura*”, conterrà le seguenti sottocartelle:

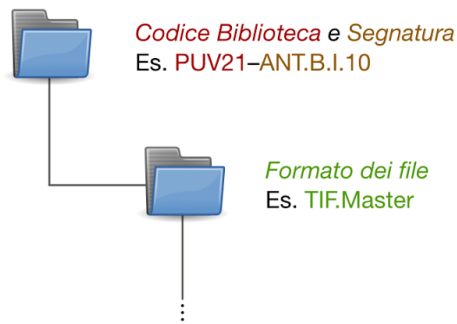
- DNG.Master, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW
- TIF.Alta.Qualita, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW; conterrà le copie fedeli all'originale
- TIF.Master, se il formato nativo prodotto dallo strumento di acquisizione è il formato TIFF
- JPG
- le sottocartelle PDF e XML, nel caso il documento sia stato sottoposto a OCR, nonché una cartella per ogni altro formato di file di testo prodotto (TXT, EPUB...)

Il nome dei file seguirà il seguente schema:

¹⁴ <https://bibliotecadigitale.cab.unipd.it/aiuto>

¹⁵ Originariamente il Sistema Bibliotecario di Ateneo per identificare le singole biblioteche aveva adottato i codici dell'anagrafica SBN (previa autenticazione, si veda: https://bibliotecadigitale.cab.unipd.it/collezioni_navigazione/Members/bibliotecari/materiali_settore_bd/gl-biblioteca-digitale/gruppo-phaidra/factory/digitalizzazione/organizzazionepolifondi_ott2018.pdf). A partire dal 2020, in concomitanza con la riorganizzazione delle biblioteche dell'Università di Padova e del cambio di software gestionale, il Sistema Bibliotecario di Ateneo ha adottato un nuovo sistema di codici (previa autenticazione, si veda: https://bibliotecadigitale.cab.unipd.it/collezioni_navigazione/Members/bibliotecari/materiali_settore_bd/gl-biblioteca-digitale/gruppo-phaidra/factory/digitalizzazione/codicifondisbaalephalma.ods). Nel caso si tratti di un progetto di polo, utilizzare il codice del polo.

“Codice Biblioteca – Segnatura – Numero Progressivo.estensione”



Esempio di struttura delle cartelle e denominazione del file:
PUV21-ANT.B.I.10\TIF.Master\PUV21-ANT.B.I.10-0001.tif

Riviste

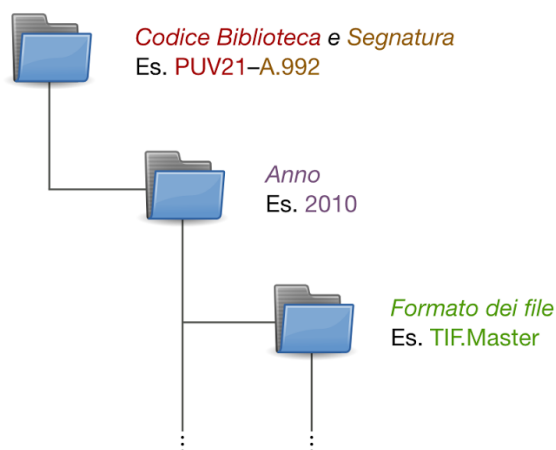
La cartella principale, denominata “*Codice Biblioteca – Segnatura*”, conterrà una sottocartella per ciascun anno della rivista, ciascuna denominata in base all'*Anno*.

All'interno dei singoli anni saranno contenute cartelle diverse per differenti tipi di file:

- DNG.Master, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW
- TIF.Alta.Qualita, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW; conterrà le copie fedeli all'originale
- TIF.Master, se il formato nativo prodotto dallo strumento di acquisizione è il formato TIFF
- JPG
- le sottocartelle PDF e XML, nel caso il documento sia stato sottoposto a OCR, nonché una cartella per ogni formato di file di testo prodotto (TXT, EPUB...)

I file saranno nominati nel seguente modo:

“Codice Biblioteca – Segnatura – Anno – Mese – Fascicolo – Numero Progressivo.estensione”



Esempio di struttura delle cartelle e denominazione del file:
 PUV21-A.992\2010\TIF.Master\PUV21-A.992-2010-12-24-0001.jpg

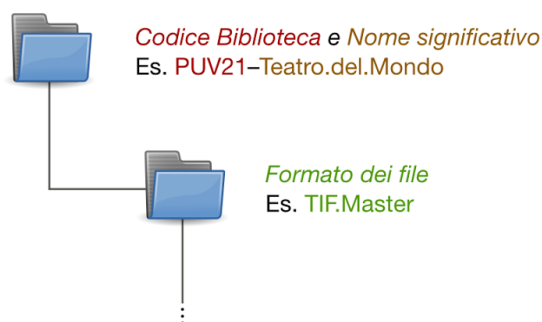
Foto, manifesti, mappe (non rilegate), pergamene e altro materiale a fogli sciolti

La cartella principale sarà denominata “*Codice Biblioteca – Nome significativo*”. Il nome significativo sarà caso per caso definito al momento della digitalizzazione; nel caso di documenti che appartengono ad un unico fondo fisico può essere semplicemente la segnatura presente nel fondo. Tale cartella conterrà le seguenti sottocartelle:

- DNG.Master, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW
- TIF.Alta.Qualita, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW; conterrà le copie fedeli all'originale
- TIF.Master, se il formato nativo prodotto dallo strumento di acquisizione è il formato TIFF
- JPG
- le sottocartelle PDF e XML, nel caso il documento sia stato sottoposto a OCR, nonché una cartella per ogni formato di file di testo prodotto (TXT, EPUB...)

Il nome dei file seguirà il seguente schema:

“Codice Biblioteca – Nome significativo – Numero Progressivo.estensione”



Esempio di struttura delle cartelle e denominazione del file:

PUV21-Teatro.del.Mondo\TIF.Master\PUV21-Teatro.del.Mondo-0001.tif

Se necessario distinguere recto da verso (es: fotografia con informazioni sul verso):

PUV21-IB.Y.1\TIF.Master-3\PUV21-IB.Y.1-0001-r.tif

PUV21-IB.Y.1\TIF.Master-3\PUV21-IB.Y.1-0001-v.tif

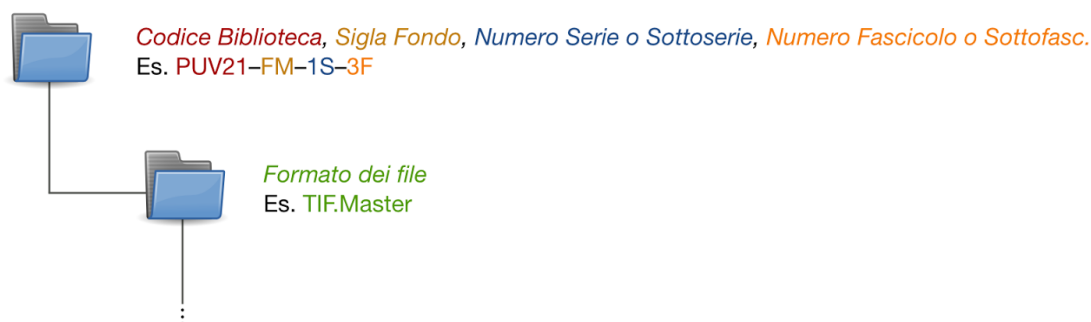
Materiale di archivio

La cartella principale, denominata “*Codice Biblioteca – Sigla Fondo – Numero Serie o Sottoserie – Numero Fascicolo o Sottofascicolo*”, conterrà le seguenti sottocartelle:

- DNG.Master, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW
- TIF.Alta.Qualita, se il formato nativo prodotto dallo strumento di acquisizione è un formato RAW; conterrà le copie fedeli all'originale
- TIF.Master, se il formato nativo prodotto dallo strumento di acquisizione è il formato TIFF
- JPG
- le sottocartelle PDF e XML, nel caso il documento sia stato sottoposto a OCR, nonché una cartella per ogni formato di file di testo prodotto (TXT, EPUB...)

Il nome dei file seguirà il seguente schema:

“*Codice Biblioteca – Sigla Fondo – Numero Serie o Sottoserie – Numero Fascicolo o Sottofascicolo – Numero Progressivo.estensione*”



Esempio di struttura delle cartelle e denominazione del file:
 PUV21-FM-1S-3F\TIF.Master\PUV21-FM-1S-3F-0001.tif

4.5 Controllo di qualità

Il controllo di qualità è finalizzato ad assicurare il rispetto delle specifiche tecniche definite a priori e la buona leggibilità a video di tutto il contenuto informativo presente negli originali; dovrebbe essere documentato ed effettuato durante l'intero processo di digitalizzazione. Oltre al controllo a video può essere utile effettuare delle prove di stampa per verificare la resa dell'immagine su carta.

La pianificazione del controllo di qualità comprende:

- la preparazione adeguata dell'ambiente (configurazione hardware, software di visualizzazione, condizioni di visualizzazione, ecc.)
- la definizione a priori delle caratteristiche di "accettabile" e "non accettabile" (ad es. se deformazioni e/o aberrazioni ottiche possano essere accettabili o no)
- le modalità di verifica (tutto il prodotto oppure un campione, tutti i file oppure solo i master, qualità visuale a schermo e a stampa, verifica strumentale, ecc.)

L'ispezione strumentale e visuale di un'immagine di solito riguarda:

- completezza della digitalizzazione e correttezza nell'ordine dei file
- correttezza dell'inquadratura e dell'esposizione
- contenimento della tolleranza cromatica
- profondità e profilo colore
- dimensione e formato digitale
- eventuale presenza di elementi che compromettono la fedeltà della riproduzione (es. riflessi luminosi)
- nomenclatura di file e cartelle

4.6 Memorizzazione e conservazione dei dati

È fondamentale mantenere nel tempo le risorse digitali create al fine di evitare di dover ripetere l'oneroso lavoro di digitalizzazione, pertanto è necessario mettere in atto procedure per assicurare

che gli oggetti digitali rimangano usabili e accessibili indipendentemente dai cambiamenti tecnologici futuri.

L'usabilità e l'accessibilità degli oggetti digitali nel tempo è garantita dal formato dei file¹⁶, dai media di archiviazione e dal digital repository. È di fondamentale importanza l'utilizzo di standard aperti.

Il Sistema Bibliotecario ha definito una procedura di archiviazione dei progetti la quale prevede che il responsabile del progetto consegni il prodotto della digitalizzazione al CAB (Centro di Ateneo per le Biblioteche) per l'archiviazione sicura presso il centro dati dell'ASIT (Area Servizi Informatici e Telematici dell'Università di Padova)¹⁷.

Per praticità di accesso, le collezioni digitali sono conservate anche su supporti di memorizzazione ottici o magnetici (*Blu-ray Disc*, dischi rimovibili), in duplice copia, una da tenere in biblioteca e l'altra presso il CAB. La copia di file da un supporto fisico ad un altro deve essere controllata con l'ausilio di software per la verifica di integrità¹⁸.

5 Archiviazione in Phaidra

L'archiviazione in Phaidra consiste nel caricare i file prodotti dalla digitalizzazione inserendo i metadati necessari per l'identificazione e la descrizione dell'oggetto digitale.

Per le modalità di archiviazione e la compilazione dei metadati, si rimanda alla *Guida all'archiviazione*¹⁹.

È possibile che gli oggetti che si vuole archiviare siano già catalogati in altri sistemi, quali il catalogo in linea o altre piattaforme, in tal caso è consigliato contattare il Servizio Aiuto del Sistema bibliotecario²⁰ per definire la procedura da adottare per l'eventuale migrazione di dati.

6 Approfondimenti

Selezione di risorse suddivise per argomento.

6.1 Progettazione

ATHENAWP3 (edited by), *Digitisation Standard Landscape for European museums, archives, libraries*, 2009

<https://phaidra.cab.unipd.it/o:6785>

Cohen, Daniel J. – Rosenzweig, R., *Digital history: a guide to gathering, preserving, and presenting the past on the web*, [2005?]

<https://chnm.gmu.edu/digitalhistory/>

¹⁶ Sui formati dei file adatti alla lunga conservazione si veda il documento [Formati dei file raccomandati per l'archiviazione a lungo termine e per la disseminazione web in Phaidra](#).

¹⁷ Previa autenticazione, si veda il documento [Archiviazione dei progetti di digitalizzazione SBA](#).

¹⁸ Ad esempio, in ambiente Windows, si può utilizzare il software *open source* [WinMerge](#).

¹⁹ <https://phaidra.cab.unipd.it/static/guida-all-archiviazione.pdf>

²⁰ <https://bibliotecadigitale.cab.unipd.it/aiuto>

International Federation of Library Associations and Institutions (IFLA), *Linee guida per pianificare la digitalizzazione di collezioni di libri rari e manoscritti*, 2015
<https://repository.ifla.org/handle/123456789/458>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida e standard*
<https://www.internetculturale.it/it/1131/linee-guida-e-standard>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Normative catalografiche, linee guida e standard*
<https://www.iccu.sbn.it/it/normative-standard/>

Istituto Centrale per la Digitalizzazione del Patrimonio Culturale – Digital Library, *Il Piano nazionale di digitalizzazione. Versione in consultazione*, 2022
<https://digitallibrary.cultura.gov.it/il-piano/>

Ministerial network for valorising activities in digitization (MINERVA), *Linee guida tecniche per i programmi di creazione di contenuti culturali digitali*, 2006
https://www.minervaeurope.org/publications/technicalguidelines_it.htm

Ministerial network for valorising activities in digitization (MINERVA), *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale*, 2004
<https://www.minervaeurope.org/publications/buonepratiche.htm>

National Information Standards Organization (NISO), *A Framework of Guidance for Building Good Digital Collections*, 2007
<https://www.niso.org/publications/framework-guidance-building-good-digital-collections>

Northeast Document Conservation Center (NDCC), *Handbook for digital projects*, 2000
<https://www.nedcc.org/assets/media/documents/dman.pdf>

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials, 2003
<https://chnm.gmu.edu/digitalhistory/links/pdf/chapter1/1.17.pdf>

6.2 Conservazione

International Federation of Library Associations and Institutions (IFLA) Core Programme, Preservation and Conservation, *Principi dell'IFLA per la cura e il trattamento dei materiali di biblioteca*, 2004
<https://repository.ifla.org/handle/123456789/1269>

Northeast Document Conservation Center (NEDCC), *NEDCC Preservation Leaflets*
<https://www.nedcc.org/free-resources/preservation-leaflets/overview>

The Library of Congress, *Preservation, Collections Care*
<https://www.loc.gov/preservation/care/>

6.3 Digitalizzazione

Association for Library Collections & Technical Services (ALCTS), *Minimum Digitization Capture Recommendations*, 2013

<https://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>

Besser, Howard (revised by S. Hubbard, D. Lenert), *Introduction to Imaging*

https://www.getty.edu/research/publications/electronic_publications/introimages/intro.html

Federal Agencies Digitization Guidelines Initiative (FADGI) - Still Image Working Group, *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*, 2016

https://www.digitizationguidelines.gov/guidelines/FADGI_Federal_Agencies_Digital_Guidelines_Initiative-2016_Final_rev1.pdf

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida e standard*

<https://www.internetculturale.it/it/1131/linee-guida-e-standard>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida per la digitalizzazione e metadati*

<https://www.iccu.sbn.it/it/normative-standard/linee-guida-per-la-digitalizzazione-e-metadati/>

Istituto Centrale per la Digitalizzazione del Patrimonio Culturale – Digital Library, *Il Piano nazionale di digitalizzazione. Versione in consultazione*, 2022

<https://digitallibrary.cultura.gov.it/il-piano/>

JISC, *JISC Digital Media* [archiviato 2016]

[https://www.webarchive.org.uk/wayback/archive/20160101151305/http://](https://www.webarchive.org.uk/wayback/archive/20160101151305/http://www.jiscdigitalmedia.ac.uk/)

www.jiscdigitalmedia.ac.uk/

Lotti, Franco – Trumpy, Giorgio, *La digitalizzazione documentaria*, CNR IFAC (Istituto di Fisica Applicata "Nello Carrara") e Fondazione Rinascimento Digitale, 2007

<https://phaidra.cab.unipd.it/o:469492>

Lotti, Franco – Lunghi, Maurizio – Trumpy, Giorgio, *Digitalizzazione di beni artistici e documentari. Manuale di procedure per un laboratorio fotografico digitale*, CNR IFAC (Istituto di Fisica Applicata "Nello Carrara") e Fondazione Rinascimento Digitale, 2009

<https://phaidra.cab.unipd.it/o:469494>

Osservatorio Tecnologico per i Beni e le Attività Culturali (OTEBAC), *Schema di capitolato per attività di digitalizzazione*, 2008

<http://www.otebac.it/index.php?it/127/capitolato-tecnico-digitalizzazione>

Research Libraries Group (RLG), *RLG Model Request for Proposal (RFP) for Digital Imaging Services*, 1997

<https://www.oclc.org/content/dam/research/activities/digimgtools/rlgmodelrfp.pdf>

University of North Texas Libraries (UNT), *Digital projects unit*
<https://library.unt.edu/digital-projects-unit/>

U.S. National Archives and Records Administration (NARA), *Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*, 2004
<https://www.archives.gov/preservation/technical/guidelines.html>

6.4 Metadati

Baca, M. (edited by), *Introduction to metadata*, 2008
https://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html

Dublin Core Metadata Initiative, *Dublin Core User Guide*
<https://www.dublincore.org/resources/userguide/>

IEEE, *Standard for Learning Object Metadata*, 2009
<https://ieeexplore.ieee.org/document/1032843>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida e standard*
<https://www.internetculturale.it/it/1131/linee-guida-e-standard>

Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), *Linee guida per la digitalizzazione e metadati*
<https://www.iccu.sbn.it/it/normative-standard/linee-guida-per-la-digitalizzazione-e-metadati/>

National Information Standards Organization (NISO), *Understanding Metadata: What is Metadata, and What is it For?*, 2017
<https://www.niso.org/publications/understanding-metadata-2017>

The Library of Congress, *Standards*
<https://www.loc.gov/librarians/standards>

-

6.5 Guide di Phaidra

Andreoli, Lorisa [et al.], *University of Vienna Metadata UWmetadata*, 2019
<https://phaidra.cab.unipd.it/static/phaidra-uwmetadata.pdf>

Bettella, Cristiana, *PHAIDRA_DC Metadata Element Set*, 2018
https://phaidra.cab.unipd.it/static/phaidra_dc-metadata-element-set.pdf

Cappellato, Linda [et al.], *Guida all'archiviazione*, 2019
<https://phaidra.cab.unipd.it/static/guida-all-archiviazione.pdf>

Drago, Gianluca (a cura di), *Formati dei file raccomandati per l'archiviazione a lungo termine*

e per la disseminazione web in Phaidra, 2019

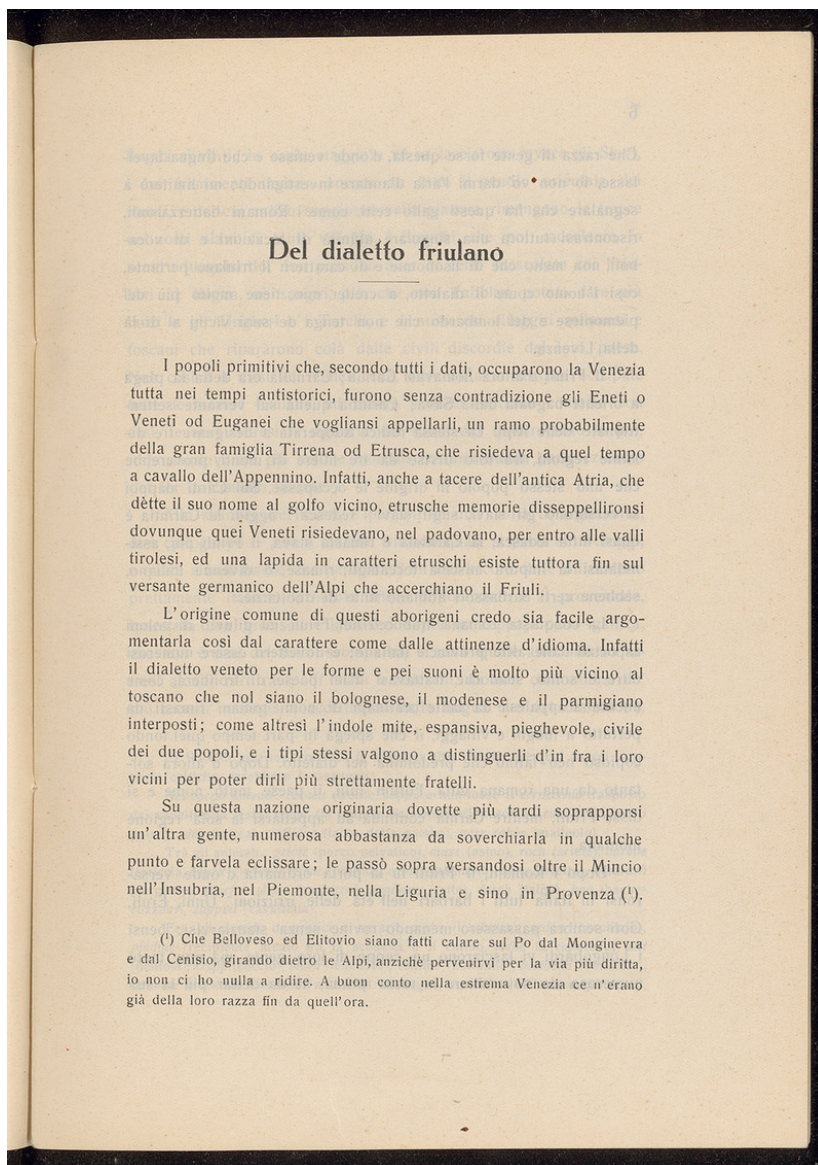
<https://phaidra.cab.unipd.it/static/formati-file-compresso.pdf>

7. Contatti

Per ulteriori informazioni sulla digitalizzazione e su Phaidra scrivere al [Servizio Aiuto](#) del Sistema bibliotecario dell'Università di Padova.

Allegato 1. Specifiche per i file XML dei testi sottoposti a OCR

Per caricare un libro sottoposto a OCR in Phaidra, i file XML devono avere lo stesso nome dei file immagine a cui si riferiscono (ad es. all'immagine pagina1.jpg deve corrispondere un file XML denominato *pagina1.xml*).



Da un'immagine come quella qui sopra (https://phaidra.cab.unipd.it/detail_object/o:83943) si dovrà ottenere un file XML così formattato:

```
<?xml version='1.0' encoding="UTF-8"?>
<ns0:ocrtext xmlns:ns0="http://phaidra.univie.ac.at/XML/book/ocrtext/V1.0">
  <ns0:page pid="o:4" abspagenum="9">
```

<ns0:ocrword word="Del" x1="542" x2="620" y1="471" y2="515">
<ns0:ocrchar char="D" x1="542" x2="581" y1="471" y2="514"/>
<ns0:ocrchar char="e" x1="585" x2="606" y1="492" y2="515"/>
<ns0:ocrchar char="l" x1="609" x2="620" y1="475" y2="515"/>
</ns0:ocrword>
<ns0:ocrword word="dialetto" x1="652" x2="821" y1="475" y2="517">
<ns0:ocrchar char="d" x1="652" x2="676" y1="475" y2="515"/>
<ns0:ocrchar char="i" x1="681" x2="690" y1="478" y2="514"/>
<ns0:ocrchar char="a" x1="694" x2="718" y1="492" y2="515"/>
<ns0:ocrchar char="l" x1="722" x2="731" y1="475" y2="514"/>
<ns0:ocrchar char="e" x1="735" x2="757" y1="492" y2="516"/>
<ns0:ocrchar char="t" x1="761" x2="774" y1="478" y2="516"/>
<ns0:ocrchar char="t" x1="778" x2="792" y1="477" y2="516"/>
<ns0:ocrchar char="o" x1="794" x2="821" y1="493" y2="517"/>
</ns0:ocrword>

[omissis]

<ns0:ocrword word="fin" x1="528" x2="557" y1="2038" y2="2058">
<ns0:ocrchar char="f" x1="528" x2="536" y1="2038" y2="2057"/>
<ns0:ocrchar char="i" x1="536" x2="543" y1="2038" y2="2058"/>
<ns0:ocrchar char="n" x1="545" x2="557" y1="2044" y2="2057"/>
</ns0:ocrword>
<ns0:ocrword word="da" x1="575" x2="604" y1="2038" y2="2059">
<ns0:ocrchar char="d" x1="575" x2="589" y1="2038" y2="2058"/>
<ns0:ocrchar char="a" x1="592" x2="604" y1="2045" y2="2059"/>
</ns0:ocrword>
<ns0:ocrword word="quell'ora." x1="621" x2="746" y1="2038" y2="2064">
<ns0:ocrchar char="q" x1="621" x2="635" y1="2045" y2="2064"/>
<ns0:ocrchar char="u" x1="639" x2="651" y1="2045" y2="2058"/>
<ns0:ocrchar char="e" x1="654" x2="666" y1="2043" y2="2058"/>
<ns0:ocrchar char="l" x1="669" x2="674" y1="2038" y2="2057"/>
<ns0:ocrchar char="l" x1="678" x2="683" y1="2038" y2="2057"/>
<ns0:ocrchar char="'" x1="686" x2="691" y1="2038" y2="2046"/>
<ns0:ocrchar char="o" x1="697" x2="711" y1="2044" y2="2058"/>
<ns0:ocrchar char="r" x1="714" x2="723" y1="2045" y2="2058"/>
<ns0:ocrchar char="a" x1="725" x2="738" y1="2045" y2="2060"/>
<ns0:ocrchar char="." x1="741" x2="746" y1="2056" y2="2060"/>
</ns0:ocrword>
</ns0:page>
</ns0:ocrtext>

Allegato 2. Scheda progetto di digitalizzazione²¹

Questa scheda indica l'insieme dei requisiti preliminari che dovrebbero essere evidenziati nei progetti di digitalizzazione finalizzati a depositare i documenti e le collezioni in Phaidra.

Informazioni sul progetto	
Titolo del progetto	
Breve descrizione della collezione/fondo	
Struttura (Dipartimento, Polo, Biblioteca, ecc.)	
Responsabile scientifico	Il responsabile scientifico (persona esperta o comitato scientifico) è chi si assume la responsabilità della selezione del materiale e che definisce la qualità dei metadati. In fase di selezione è coadiuvato dal responsabile di progetto, in particolare per l'esame dei materiali e degli aspetti legali.
Nominativo	
Telefono	
Email	
Responsabile di progetto	Il responsabile di progetto coopera con il responsabile scientifico, supporta il responsabile scientifico nell'analisi degli aspetti legali, coordina le attività relative alla digitalizzazione e garantisce la qualità dei metadati.
Nominativo	
Telefono	
Email	
Referente tecnico	È il referente per le attività tecnico-operative che collabora con il gruppo Phaidra, a sua volta garante dell'assistenza tecnica.
Nominativo	
Telefono	
Email	

²¹ La Scheda progetto digitalizzazione è scaricabile, previa autenticazione, all'indirizzo https://bibliotecadigitale.cab.unipd.it/collezioni_navigazione/Members/bibliotecari/materiali_settore_bd/gl-biblioteca-digitale/gruppo-phaidra/factory/digitalizzazione/elenco-dei-documenti-relativi-alla-digitalizzazione

Breve descrizione delle fasi del progetto e degli attori coinvolti	
Durata del progetto	
Informazioni sui documenti originali (libro, annata di rivista, atlante, mappa, fotografia, ecc.)	
Datazione	da _____ a _____
Tipologia	Quantità stimata
<input type="checkbox"/> testo a stampa	
<input type="checkbox"/> testo manoscritto	
<input type="checkbox"/> musica a stampa e manoscritta	
<input type="checkbox"/> mappa	
<input type="checkbox"/> manifesto	
<input type="checkbox"/> cartolina	
<input type="checkbox"/> disegno	
<input type="checkbox"/> dipinto	
<input type="checkbox"/> stampa (incisione, acquaforte, ecc.)	
<input type="checkbox"/> pergamena	
<input type="checkbox"/> negativo b/n	
<input type="checkbox"/> negativo col.	
<input type="checkbox"/> fotografia b/n	
<input type="checkbox"/> fotografia col.	
<input type="checkbox"/> diapositiva b/n	
<input type="checkbox"/> diapositiva col.	
<input type="checkbox"/> altro (specificare) _____	
Come si presentano i documenti	<input type="checkbox"/> fogli sciolti <input type="checkbox"/> fogli arrotolati <input type="checkbox"/> rilegato <input type="checkbox"/> album <input type="checkbox"/> montaggio su cartone o altro materiale <input type="checkbox"/> montaggio su cornice <input type="checkbox"/> buste <input type="checkbox"/> cartelle

	<input type="checkbox"/> scatole <input type="checkbox"/> altro (specificare) <hr/>
Dimensioni dei documenti	< di A4 <hr/> A4 <hr/> A3 <hr/> A2 <hr/> A1 <hr/> A0 <hr/> > di A0 <hr/> altro (specificare) <hr/>
Numero totale dei documenti	
Informazioni sugli oggetti digitali (file prodotti)	
Numero stimato di oggetti digitali	
Destinazione d'uso degli oggetti digitali	<input type="checkbox"/> accesso web aperto * <input type="checkbox"/> accesso web ristretto * <input type="checkbox"/> accesso in rete locale <input type="checkbox"/> CD-ROM o DVD <input type="checkbox"/> stampa <input type="checkbox"/> altro (specificare) <hr/>

	<p>* Si ricorda che gli oggetti digitali inseriti in Phaidra potranno essere ad accesso aperto (tutti gli utenti potranno visualizzare la preview, i metadati e il file) ovvero ad accesso ristretto (visione dei metadati).</p>
Verifiche preliminari	
Provenienza dei documenti	<p> <input type="checkbox"/> acquisizione <input type="checkbox"/> donazione <input type="checkbox"/> non so <input type="checkbox"/> altro (specificare) </p> <hr/> <hr/> <hr/>
È stata fatta una selezione dei documenti?	<p> <input type="checkbox"/> sì <input type="checkbox"/> in parte <input type="checkbox"/> no </p> <p>Se sì, quali sono i criteri di selezione?</p> <p> <input type="checkbox"/> valore storico e culturale <input type="checkbox"/> unicità e rarità <input type="checkbox"/> alta richiesta <input type="checkbox"/> materiale senza vincoli legali o con permessi di digitalizzazione ottenuti <input type="checkbox"/> accesso ristretto dovuto allo stato di conservazione, al valore, alla localizzazione <input type="checkbox"/> valore aggiunto tramite l'accesso in linea, la creazione di collezioni virtuali, l'aumento di interesse di ricerca per materiale poco o non conosciuto, ecc. <input type="checkbox"/> altro (specificare) </p> <hr/> <hr/>
È stato fatto un censimento?	<p> <input type="checkbox"/> sì <input type="checkbox"/> in parte <input type="checkbox"/> no </p>
Esiste una versione digitalizzata?	<p> <input type="checkbox"/> sì <input type="checkbox"/> no </p> <p>Se no, quali enti, siti, cataloghi, ecc. sono stati controllati?</p> <hr/>

	<hr/> <hr/> <hr/>
Ci sono vincoli legali (diritto d'autore, tutela della riservatezza, diritti del donatore, ecc.)?	<input type="checkbox"/> sì <input type="checkbox"/> no <input type="checkbox"/> non so Eventuali informazioni aggiuntive: <hr/> <hr/>
I documenti sono descritti/catalogati?	<input type="checkbox"/> sì, tutti <input type="checkbox"/> sì, una parte <input type="checkbox"/> no <input type="checkbox"/> non so Se sì, in che modo? <input type="checkbox"/> elenco cartaceo <input type="checkbox"/> elenco elettronico <input type="checkbox"/> catalogazione cartacea <input type="checkbox"/> catalogazione elettronica <input type="checkbox"/> inventariazione archivistica cartacea <input type="checkbox"/> inventariazione archivistica elettronica <input type="checkbox"/> altro (specificare) <hr/> <hr/>
Nel caso di testo a stampa, s'intende applicare l'OCR (Optical Character Recognition)	<input type="checkbox"/> sì <input type="checkbox"/> no <input type="checkbox"/> non so
Nel caso di testo manoscritto, s'intende trascrivere i documenti?	<input type="checkbox"/> sì, tutti <input type="checkbox"/> sì, una parte <input type="checkbox"/> no <input type="checkbox"/> non so
Stima dei costi del progetto	Se la digitalizzazione è <i>in-house</i> , indicare: • costo attrezzatura specificando il tipo di strumentazione <hr/> <hr/>

	<ul style="list-style-type: none"> • costo operatore <hr/> <p>Se la digitalizzazione è in <i>outsourcing</i>, indicare:</p> <ul style="list-style-type: none"> • costo unitario <hr/> <ul style="list-style-type: none"> • costo complessivo <hr/>
Note	
Scheda compilata da	
Data di compilazione	

I sottoscritti sono consapevoli di dover operare nel rispetto della normativa vigente sul diritto d'autore.

I documenti di questo progetto, soggetti alla normativa in vigore sul diritto d'autore, rientrano in una delle seguenti condizioni (barrare una o più opzioni):

- il detentore dei diritti è l'Università di Padova
- i detentori dei diritti hanno concesso all'Università di Padova il diritto a utilizzarli
- sono di pubblico dominio

Firma del Responsabile scientifico _____

Firma del Responsabile di progetto _____