

**Procedura: Servizio di digitalizzazione di alcuni documenti, di diverse tipologie e dimensioni, appartenenti alla Biblioteca dell'Orto Botanico dell'Università di Padova, tramite Mercato Elettronico della Pubblica Amministrazione (MePA).
C.I.G.: Z9D1FC1138.**

CAPITOLATO - PARTE TECNICA

Oggetto dell'appalto

Il Centro di Ateneo per le Biblioteche (CAB), in qualità di coordinatore dei progetti di digitalizzazione, intende affidare la digitalizzazione di libri, opuscoli, lettere, manoscritti e disegni dell'archivio del Prefetto dell'Orto Botanico Pier Andrea Saccardo, nell'ambito del progetto *Strategie innovative per la valorizzazione e la diffusione della conoscenza di collezioni botaniche e bibliografiche ad esse correlate*.

Documenti da digitalizzare

I documenti sono conservati presso la Biblioteca dell'Orto Botanico di Padova.

Documenti da digitalizzare:

- *Fungi italici*, libro di fine '800 in formato A3, di p. 774
- *L'erbario micologico di P.A. Saccardo: catalogo*, libro di inizio '900 in formato A4, di p. 354. Di questo libro si chiede trascrizione del testo in OCR
- Vari fogli manoscritti e dattiloscritti di fine '800 – inizio '900, per lo più sciolti, di agevole manipolazione, non particolarmente delicati

Formato dei documenti da digitalizzare (libri + fogli di archivio):

- A4: 5.182
- A3: 1.574

N. totale dei file immagine da produrre: 6.756 che corrispondono ai documenti elencati nell'*Allegato 1*.

Nel caso in cui con la digitalizzazione dei documenti elencati non venisse raggiunta la spesa di € 10.000,00 IVA e oneri di sicurezza inclusi, il CAB selezionerà altri documenti fino ad arrivare a tale cifra.

Generalità

La digitalizzazione mira alla creazione di file destinati alla lunga conservazione, *file master*, e di file frutto di elaborazioni successive, *file derivati*, destinati alla fruizione da parte degli utenti, in genere via Web.

Sarà compito dell'aggiudicatario procurare il personale qualificato e tutta l'attrezzatura appropriata per il compimento dell'attività di digitalizzazione, nel pieno rispetto della normativa vigente in tema di sicurezza del lavoro.

Il CAB è esente da ogni responsabilità per danni a persone e/o cose provocati durante le operazioni di digitalizzazione. Per ogni effetto, l'aggiudicatario assume la piena responsabilità di danni a persone, opere e/o cose conseguenti alle attività svolte.

Sarà a carico dell'aggiudicatario la stipula di una assicurazione per coprire i rischi di lavorazione.

Tenendo conto del pregio e della delicatezza degli originali, le modalità di trattamento (trasporto, illuminazione, manipolazione) dovranno essere tali da non arrecare loro danno.

La digitalizzazione dovrà tenersi presso l'edificio ove ha sede la Biblioteca dell'Orto Botanico di Padova – Via Orto Botanico, 15, 35123 Padova – dove dovranno essere montate le attrezzature dell'aggiudicatario, e verrà fatta in un orario concordato, all'interno dell'orario di ufficio (da lunedì a venerdì, ore 9.00-17.00) **a partire dalla data di stipula del contratto**. L'elaborazione digitale delle immagini potrà essere svolta in altra sede.

I prodotti della digitalizzazione dovranno essere consegnati presso la Biblioteca dell'Orto Botanico – via Orto Botanico 15, 35123 Padova – **entro 30 giorni lavorativi dall'inizio dell'attività**.

Specifiche

Specifiche tecniche dell'attrezzatura

Il sistema di acquisizione (sorgente di luce, ottica, sensore, software di acquisizione e di calibrazione) deve garantire la qualità di immagine richiesta al punto "Caratteristiche dei file prodotti" e non danneggiare i documenti originali. Per non danneggiare i documenti il sistema di acquisizione da usare può essere solo lo scanner planetario o la macchina fotografica digitale.

Il sistema di illuminazione deve essere a luce fredda senza emissione di UV e IR.

Caratteristiche dei file prodotti

File master

- È archiviato così come lo restituisce lo strumento di acquisizione.
- **Area di ripresa:** il documento deve essere ripreso nella sua interezza.

1) **Libri e altri documenti rilegati:** attorno al documento, si lascia un bordo nero di 3-4 millimetri che permetta di leggere il contorno del documento. Tuttavia, dal momento che i libri verranno visualizzati in una applicazione che consente di sfogliarne le pagine, il lato della pagina verso l'interno del libro non dovrà avere bordo, o dovrà avere un bordo minimo (vedi figura a fianco).



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 presenza di lacerazioni, di fori di tarlo, di ossidazione degli inchiostri, di carta che lascia trasparire il contenuto sottostante, le carte devono essere mascherate con carta bianca in modo da evitare di riprendere il contenuto sottostante.

2) **Manoscritti, lettere, disegni e altri documenti non rilegati:** attorno al documento, si lascia un bordo nero di 3-4 millimetri che permetta di leggere il contorno del documento. Verrà prodotto un file immagine per ciascun recto e ciascun verso, con l'esclusione delle pagine bianche.

Ogni documento deve essere digitalizzato affiancato da scala cromatica, scala di grigi e scala metrica, poste all'esterno dell'immagine riprodotta e all'interno dell'inquadratura complessiva. Nel caso dei libri è sufficiente posizionare le scale una sola volta su una carta o pagina (che verrà digitalizzata 2 volte, una con le scale e una senza).

Specifiche tecniche per i file master:

- Il file deve incorporare un profilo di colore “Adobe RGB (1998)” con profondità di 24 bit (8 per canale)
- La risoluzione ottica non deve essere inferiore a 600 dpi fino al formato A4, e ai 400 dpi per formati più grandi di A4
- Deve essere in formato TIFF 6.0, non compresso

File derivati

Dai file derivati dovranno essere rimosse le scale cromatiche, di grigi e metriche.

I file derivati dovranno essere bilanciati per 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.

Tale bilanciamento deve mirare alla riproduzione fedele delle caratteristiche cromatiche dell'originale, non ad un arbitrario miglioramento estetico.

Dovranno essere prodotti tutti i seguenti tipi di file derivati:

- **TIFF 6.0**, non compresso, con profilo colore “Adobe RGB (1998)” e profondità di 24 bit (8 bit per canale), alla stessa risoluzione del master, riscalato ad una dimensione del lato lungo di 2400 pixel.
- **JPEG di buona qualità**, con profilo colore “sRGB IEC-61966-2.1” e profondità di 24 bit (8 bit per canale), a 300 dpi, compresso alla qualità massima (100%), riscalato ad una dimensione del lato lungo di 2400 pixel.
- **JPEG di bassa qualità**, con profilo colore “sRGB IEC-61966-2.1” e profondità di 24 bit (8 bit per canale), a 150 dpi, compresso ad una qualità compresa tra 90% e 100%, riscalato ad una dimensione del lato lungo di 1500 pixel.

OCR

Per il libro di cui si chiede riconoscimento del testo in OCR, dovrà essere fornito un PDF con testo ricercabile, corredato da un file XML formattato come da modello nell'*All'egato 2* e un file di testo semplice contenente il solo contenuto testuale del libro.

L'OCR dovrà essere fatto sul file derivato JPG di bassa qualità (150 dpi, dimensione del lato lungo 1500 px).

Controlli della qualità

L'aggiudicatario è tenuto a garantire la competenza degli operatori impiegati e a effettuare tutti gli opportuni controlli di qualità sull'attività di digitalizzazione.

Il sistema di controllo della qualità sarà finalizzato ad assicurare l'accuratezza dei requisiti richiesti e la buona leggibilità a video di tutto il contenuto informativo presente negli originali e consisterà in:

- creazione di un prototipo del prodotto sulla base di un campione significativo dei materiali forniti dalle biblioteche; l'approvazione formale del prototipo costituirà riferimento di qualità per il successivo processo di scansione/ripresa
- verifiche dei file prodotti a cura del CAB che avrà facoltà di chiedere la ripetizione delle scansioni/ripresе difettose senza che l'aggiudicatario possa pretendere compensi oltre il corrispettivo dovuto.

Nomenclatura dei file

I file dovranno essere denominati seguendo le indicazioni fornite in *Allegato 3*.

Modalità di consegna del prodotto

Al termine dell'attività, la collezione delle immagini consistente in cartelle e file sarà consegnata su dischi rigidi esterni: due dischi – di marca o serie differente – che saranno consegnati al CAB, presso la Biblioteca dell'Orto Botanico in via Orto Botanico 15, 35123 Padova.

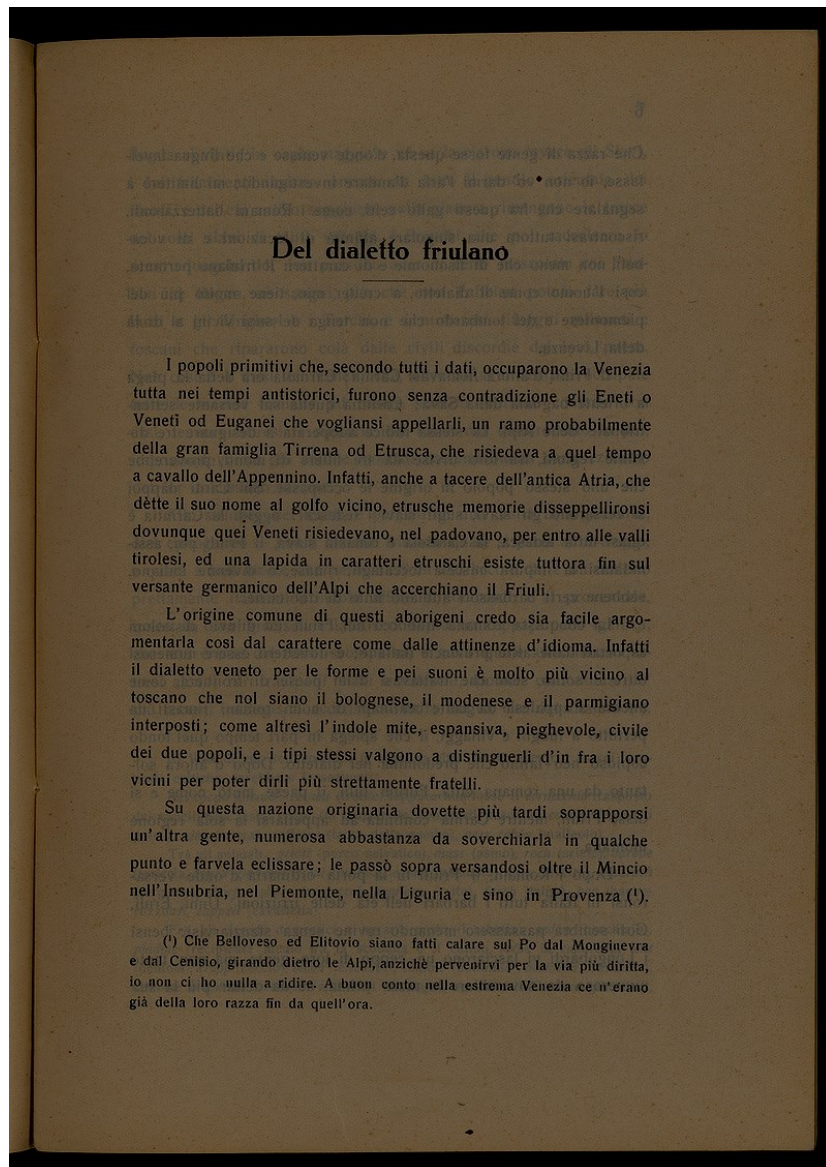
L'aggiudicatario dovrà indicare marca, modello e durata della garanzia dei supporti.

L'aggiudicatario è comunque tenuto a conservare i file master per un periodo di tempo sufficiente a consentire il completamento del controllo di qualità da parte del CAB.

L'aggiudicatario consegnerà anche un breve rapporto descrittivo dell'attività svolta, delle scelte tecniche operate e dei contenuti dei dischi rigidi esterni.

Allegato 2. Specifiche per i file XML dei testi da sottoporre a OCR

I file XML devono avere lo stesso nome dei file immagine a cui si riferiscono (pagina1.jpg -> pagina1.xml).



Da un'immagine come quella qui sopra (l'esempio è tratto: 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 3 - Nomenclatura dei file

In generale il nome di ciascun file sarà una stringa di caratteri composta da più parti, avente in essa le informazioni necessarie ad identificare in maniera univoca il documento del progetto a cui l'immagine si riferisce. I nomi dei file saranno completati a cura dell'aggiudicatario con l'estensione adeguata (tif, jpg, pdf, xml).

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

L'organizzazione delle cartelle principali è la seguente: una cartella per ogni libro, una cartella per ogni faldone del materiale d'archivio della Biblioteca.

Per la nomenclatura dei file si dovrà seguire la struttura qui descritta e ulteriormente concordata in fase di realizzazione del prototipo.

La nomenclatura delle cartelle e dei file è una stringa di campi (Codice Biblioteca, Collocazione...) separati da trattino (-). Laddove la collocazione contenga trattini (-), spazi o caratteri speciali, questi vengono rimpiazzati da un punto (.). Si veda l'*Allegato 1* per il dettaglio sulla denominazione delle cartelle.

Coperta anteriore e posteriore dei libri vanno nominate in modo che si presentino nello stesso ordine che hanno nel documento fisico. Eventuali dorso o altre parti del documento originale (tagli, dettagli della legatura...) devono presentarsi alla fine.

L'immagine che include le scale di colore e di grigio ed il righello, deve essere nominata in modo che sia l'ultimo file della cartella e alla numerazione progressiva del file va aggiunta una "-c".

La cartella principale, denominata "*Codice Biblioteca – Collocazione*", conterrà le seguenti sottocartelle: *TIF.Master*, *TIF.Derivato*, *JPG300*, *JPG150*, *PDF*, *XML* (quest'ultima solo per il libro di cui si chiede OCR).

Per i **libri**, i file avranno lo stesso nome della cartella principale seguito da trattino e numero progressivo di 4 cifre + estensione.

"Codice Biblioteca – Collocazione – Numero progressivo.estensione"

Esempio:

PUV46-A.H.P.45\TIF.Master\PUV46-A.H.P.45-0001.tif

Per i **documenti di archivio della Biblioteca**, i file avranno lo stesso nome della cartella principale seguito da trattino, seguito dal codice documento, seguito da numero progressivo di 4 cifre + estensione.

"Codice Biblioteca – Collocazione – Numero di faldone – Codice documento – Numero progressivo.estensione"

Esempio:

PUV46-Ar.B.32\TIF.Master\PUV46-Ar.B.32-025-0001.tif