

Adresse  
AP  
Str.

**PLZ Ort**

Ort, den

## **Image-Digitalisierung Bitte um ein Angebot**

Sehr geehrte Damen und Herren,

Zweck des Anschreibens

### **1 Allgemeine Hinweise**

Art der Einrichtung, Voraussetzung, Abstimmungsgespräche, Ansprechpartner benennen, REferenzen

### **2 Image-Digitalisierung**

#### **2.1 Projektzeitraum und -umfang**

Projektlaufzeit:

Volumen: Titel mit ca. Seiten

#### **2.2 Beschreibung der Digitalisierungsvorlagen**

Beschreibung

Die Vorlagen (Bücher) lassen sich 180°, ggf. jedoch nur 60 - 90° öffnen. Falls sich hieraus verschiedene Kosten ergeben, bitten wir um getrennte Angaben:

- Kosten je Digitalisat (Einzelseite) bei einem Öffnungswinkel 180°
- Kosten je Digitalisat (Einzelseite) bei einem Öffnungswinkel kleiner 180°

#### **2.3 Leistungsumfang**

##### **2.3.1 Transport und Versicherung**

Kollegium:  
Ute Frevert  
Gerd Gigerenzer  
Ulman Lindenberger

Lentzeallee 94 D-14195 Berlin (Dahlem)  
Telefon: 030-824 06 229  
Telefax: 030-824 9939  
e-mail: flitner@mpib-berlin.mpg.de  
<http://www.mpib-berlin.mpg.de>

1. Transport der zu bearbeitenden Bände zwischen xxx und Dienstleistungsunternehmen.
2. Versicherung der zu digitalisierenden Bände (Haftung für Schäden innerhalb der Betriebsräume und während des Transports bis zu einer Höhe von € xxx,- je Band).

### **2.3.2 Testlauf**

Um einen Eindruck von genutzter Hardware, Aufwand und Qualität der Scans zu erhalten, möchten wir mit einem repräsentativen Bestand testen.

### **2.3.3 Digitalisierung**

1. Erzeugung von Farbdigitalisaten mit mind. 400 dpi Auflösung und 24 Bit-Farbtiefe innerhalb eines Monats nach Vertragsabschluss.
2. Verwendung von Farbtafel und Zentimetermaß bei der ersten Aufnahme je Band (Anbringen auf U2).
3. Herstellung von digital master-Dateien im Format Baseline-TIFF gemäß DFG-Digitalisierungsrichtlinien (s. a. Anlage: Image-Format: Baseline TIFF). Jede Seite wird einzeln abgespeichert und benannt.

### **2.3.4 Bildbearbeitung**

Ggf. erfolgt eine Bildbearbeitung:

1. ggf. Vereinzelung von Doppelseiten
2. ggf. Kontrast-, Helligkeits- und Schärfenkorrektur
3. ggf. gerades Ausrichten des Satzspiegels
4. ggf. Beseitigung von Doppelaufnahmen
5. ggf. Korrektur der Bild-Reihenfolge nach den Vorgaben des MPIB
6. Grundsätzlich gilt: keine Randentfernung – Es muss gewährleistet sein, dass die Größe des Buches als Eindruck erhalten bleibt, d. h., dass ein schmaler schwarzer Rand erhalten bleiben muss.

### **2.3.5 Herstellung von Derivaten**

Es sollen Derivate im jpg-Format generiert werden. Jede Seite wird einzeln abgespeichert und benannt im JPEG-Format (24 Bit-Farbtiefe).

### **2.3.6 Datenauslieferung**

1. Verzeichnisstruktur: Speicherung der Bilddateien jedes Bandes in einem separaten Verzeichnis; Benennung des Verzeichnisses nach Vorgabe. Ordnerstruktur s. Anhang
2. Dateibenennung: Jeder Dateiname setzt sich wie folgt zusammen:...
3. Bereitstellung der digital master-Dateien (TIFF) und JPG-Dateien (Auflösung: 24 bit) via ftp/Festplatte...
4. Übermittlung der digital master-Dateien (TIFF) an eine Archivierungsstelle

## **2.4 Qualitätsmanagement**

### **2.4.1 Erstellung von Digitalisierungs-Protokollen**

Das Digitalisierungsprotokoll zu jedem Band enthält: Verzeichnisname, Titel des Werkes, ggf. Band, Hinweis auf gelöschte (Doppel-) Bilder, Korrekturen der Bildreihenfolge, sonstige Bemerkungen.

### **2.4.2 Qualifizierte Prüfung der Digitalisierung**

1. Wurde die Vorlage vollständig digitalisiert? Entspricht die Zahl der Scans den Angaben des MPIB über die Seitenzahl? Bei Diskrepanzen ist mit dem MPIB Rücksprache zu halten.
2. Ist die Bildabfolge korrekt?
3. Sind alle TIFF- und jpg-Dateien lesbar?
4. Entspricht die Bildwiedergabe den Qualitätsanforderungen, d. h., ist die Schrift auf der vollen Fläche der Seite klar lesbar, der Satzspiegel gerade ausgerichtet?
5. Ist der Verzeichnisname korrekt?

### **2.4.3 Abnahme und Freigabe zur Datenvernichtung**

Eine endgültige Abnahme der Daten erfolgt durch xxx. Für nötige Korrekturen trägt der Auftragnehmer die Kosten, es sei denn, sie resultieren aus fehlerhaften Angaben und Anweisungen des Auftraggebers. Die Daten sind auf den Servern des Auftragnehmers so lange aufzubewahren, bis der Auftraggeber den Auftrag zum Löschen erteilt.

### **2.4.4 Regressforderung**

Für im Digitalisierungsprozess durch den Dienstleister beschädigte Bücher können Regressforderungen gestellt werden.

## **2.5 Angebot**

Bitte erstellen Sie ein Angebot über Kosten je Digitalisat (Einzelseite) Baseline-TIFF, ggf. differenziert:

- Kosten je Digitalisat (Einzelseite) bei einem Öffnungswinkel 180°
- Kosten je Digitalisat (Einzelseite) bei einem Öffnungswinkel kleiner 180°

## **2.6 Zusatzinformationen**

Bitte fügen Sie dem Angebot Informationen zu folgenden Punkten bei:

- a. Haben Sie bereits Erfahrungen in der Zusammenarbeit mit Einrichtungen unserer Art?
- b. Welche Hardware kommt bei der Durchführung des Projekts voraussichtlich zum Einsatz? (Hersteller / Modell)
- c. Über welche Kapazität verfügt Ihre Internet-Verbindung für die Online-Übermittlung der Daten?
- d. Welche Speicherkapazität steht auf Ihren Servern für die Daten des Projekts bereit?

- e. Wie groß schätzen Sie das Speichervolumen der zu produzierenden Daten?
- f. Es wird für jedes Buch einen Laufzettel geben, der immer direkt am Buch bleibt, um evtl. Anmerkungen und den Stand der Digitalisierung zu notieren.
- g. Ihr Angebot muss sich auf die in der Angebotseinholung gemachten Angaben beziehen.

### **3 Erfassung des Volltextes**

Die digitalisierten Bände werden via OCR volltexterfasst und in eine xml-Struktur nach Vorgaben des xxx gebracht (angelehnt an TEI).

#### **3.1 Leistungsumfang**

1. Erfassung des Volltextes als dirty OCR. Eine Korrektur der Ergebnisse erfolgt nicht.
2. Der gewonnene Volltext ist in eine xml-Struktur gemäß den Vorgaben zu bringen (Anlehnung an TEI). Für die korrekte und nachvollziehbare Ermittlung der Zeichenzahl hat der Auftragnehmer Sorge zu tragen.

#### **3.2 Qualitätsmanagement**

1. Vollständigkeit der Daten
2. Qualifizierte Prüfung der xml-Struktur

#### **3.3 Angebot**

Bitte erstellen Sie ein Angebot über die Kosten für **je 1.000 erfasste Zeichen**.

Wir bitten um Ihr Angebot bis zum xxx . Für Rückfragen stehen wir gerne zur Verfügung. Ansprechpartner ist Name, Telefonnummer, Mail-Adresse.

Mit freundlichen Grüßen

## 4 Anlage

### 4.1 Image-Format "Baseline TIFF"

#### 4.1.1 Vorbemerkung

"Baseline" TIFFs sind TIF-Dateien, die sich auf die Verwendung eines *Subsets* der innerhalb einer TIF Datei insgesamt verwendbarer "Tags" beschränken.

Eine genaue Beschreibung des Konzepts "Baseline TIFF" bildet Kapitel 2 der Formatspezifikation, die derzeit in Version 6, final revision, June 3rd 1992, vorliegt. Die gültige Version der Spezifikation ist erhältlich über <http://partners.adobe.com/public/developer/tiff/index.html>. Die folgende Darstellung fasst die wesentlichen Punkte zusammen.

#### 4.1.2 Was ist ein Baseline TIFF?

##### 4.1.2.1 Strenge technische Definition

Eine Baseline TIFF Datei ist eine TIFF Datei, die folgende in der vollständigen TIFF Spezifikation zulässigen Features **vermeidet**:

1. Mehrere voneinander unabhängige Bilder in *einer* Datei. (TIFF Spezifikation p.16)
2. "Private fields", also Tags oder Werte, die nicht im Standard spezifiziert sind, auch dann, wenn diese Tags registriert wurden. (TIFF Spezifikation p. 8)
3. CCITT Group 3 and 4 Kompression. (Die von der CCITT Spezifikation abgeleitete 1D Kompression *IST* Teil der baseline Spezifikation.) (TIFF Spezifikation p. 49 ff.)
4. Document Storage and Retrieval Fields. (TIFF Spezifikation p. 55 ff.)
5. LZW Kompression. (TIFF Spezifikation p. 57 ff.)
6. Differencing Predictor für die Kompression. (TIFF Spezifikation p. 64 f.)
7. CMYK Bilder. (TIFF Spezifikation p. 69 ff.)
8. Halftone hints. (TIFF Spezifikation p. 72 ff.)
9. Associated Alpha Handling. (TIFF Spezifikation p. 77 ff.)
10. Non-standard data sample formats. (TIFF Spezifikation p. 80 f.)
11. Colorimetry für RGB Bilder. (TIFF Spezifikation p. 82 ff.)
12. YCbCr Bilder. (TIFF Spezifikation p. 89 ff.)
13. JPEG Kompression innerhalb von TIFF Bildern. (TIFF Spezifikation p. 95 ff.)
14. CIE L\*a\*b Bilder. (TIFF Spezifikation p. 110 ff.)

##### 4.1.2.2 Modifikationen im Sinne der Langzeitarchivierung

Nur für TIF Dateien, die die oben genannten Features **vermeiden**, ist garantiert, dass *jeder* existierende TIFF Reader sie auch wirklich verarbeiten kann. Die Implementationen der TIFF

Spezifikationen haben bestimmte optionale Features jedoch so häufig verfügbar gemacht, dass eine Reihe von Verletzungen der oben genannten Regeln bewusst hingenommen, in zwei der folgenden Fälle sogar empfohlen werden.

- a. Im Interesse der langfristigen Verwendbarkeit von Bildern liegt es, **Metadaten** in diese einzufügen. Der dafür empfohlene internationale **EXIF Standard** (<http://www.exif.org>) ist mittlerweile so weit verbreitet, dass die Einführung einer Exif-Beschreibung in eine TIF Datei befürwortet wird. Dafür wird innerhalb von TIF das private field "EXIF IFD pointer" (34665(0x8769) ) reserviert.
- b. Im Interesse der langfristigen Verwendbarkeit von Bildern liegt es ferner, ein **ICC Farbprofil** in der Datei zu speichern. Der dafür festgelegte Standard (vgl. [http://www.color.org/icc\\_specs2.xalter](http://www.color.org/icc_specs2.xalter)) legt dafür das TIF private field "ICC Profile" (34675 (0x8773) ) fest.

#### 4.1.3 Wie erkennt man Baseline TIFFs?

Alle professionellen Bildverarbeitungsprogramme sollten den einen oder anderen Mechanismus anbieten, um die in einer TIF Datei enthaltenen Tags darzustellen. Dies wird in der Regel entweder durch den Standardnamen des Tags, durch dessen Übersetzung in eine andere Sprache, seinen (dezimalen) numerischen Wert oder seinen hexadezimalen Wert erfolgen.

*Eine TIF Datei, die dabei ausschließlich Tags aus den ersten drei folgenden Listen, und nicht mehr als genau ein Bild enthält, ist im Sinne dieser Empfehlungen eine Baseline TIF Datei. Enthält sie ein in der vierten Teilliste enthaltenes Tag, ist sie ein Risiko für die Langzeitspeicherung.*

- I. Eine Teilmenge (in Abhängigkeit von dem Bildtyp – Monochrom, Grauwert, Indexed, RGB) der folgenden Tags **muss** angegeben werden:

(Standardname: Dezimalwert, Hexwert)

ImageWidth: 256, 0x100.

ImageLength: 257, 0x101.

BitsPerSample: 258, 0x102.

Compression: 259, 0x103.

Jedoch nur mit den Werten: (Standardname: Code) Uncompressed: 1; CCITT 1D: 2;

PackBits: 32773

PhotometricInterpretation: 262, 0x106.

Jedoch nur mit den Werten: (Standardname: Code) WhiteIsZero: 0; BlackIsZero: 1;

RGB: 2; RGBPalette: 3

StripOffsets: 273, 0x111.

SamplesPerPixel: 277, 0x115.

RowsPerStrip: 278, 0x116.

StripByteCounts: 279, 0x117.  
XResolution: 282, 0x11A.  
YResolution: 283, 0x11B.  
ResolutionUnit: 296, 0x128.  
ColorMap: 320, 0x140.

II. Die folgenden Tags müssen von allen Baseline-Readern verarbeitet werden. Ihre Präsenz ist für die Verarbeitung jedoch nur nötig, wenn von Standardvorgaben abgegangen wird.

(Standardname: Dezimalwert, Hexwert)

NewSubfileType: 254, 0xFE.  
SubfileType: 255, 0xFF.  
Threshholding: 263, 0x107.  
CellWidth: 264, 0x108.  
CellLength: 265, 0x109.  
FillOrder: 266, 0x10A.  
ImageDescription: 270, 0x10E.  
Make: 271, 0x10F.  
Model: 272, 0x110.  
Orientation: 274, 0x112.  
MinSampleValue: 280, 0x118.  
MaxSampleValue: 281, 0x119.  
PlanarConfiguration: 284, 0x11C.  
FreeOffsets: 288, 0x120.  
FreeByteCounts: 289, 0x121.  
GrayResponseUnit: 290, 0x122.  
GrayResponseCurve: 291, 0x123.  
Software: 305, 0x131.  
DateTime: 306, 0x132.  
Artist: 315, 0x13B.  
HostComputer: 316, 0x13C.  
ExtraSamples: 338, 0x152.  
Copyright: 33432, 0x8298.

III. Die folgenden Tags sind nicht in der Baseline Definition enthalten. Sie sind jedoch so weitverbreitet / so nützlich, dass ihre Verwendung unbedenklich ist. (S. oben.)



(Standardname: Dezimalwert, Hexwert)

Compression: 259, 0x103.

Auch mit den Werten: (Standardname: Code) Group3Fax: 3; Group4Fax: 4;

EXIF IFD pointer: 34665, 0x8769.

ICC Profile: 34675, 0x8773.

IV. Die folgenden Tags sind Teile von Erweiterungen, deren Unterstützung explizit optional ist. Bei ihrer Verwendung ist also damit zu rechnen, dass viele TIFF Reader die Bilddaten und / oder im Bild gespeicherte Metadaten **nicht** oder nur eingeschränkt verarbeiten können.

(Standardname: Dezimalwert, Hexwert)

PageName: 285, 0x11D.

XPosition: 286, 0x11E.

YPosition: 287, 0x11F.

PageNumber: 297, 0x129.

TransferFunction: 301, 0x12D.

Predictor: 317, 0x13D.

WhitePoint: 318, 0x13E.

PrimaryChromaticities: 319, 0x13F.

HalftoneHints: 321, 0x141.

TileWidth: 322, 0x142.

TileLength: 323, 0x143.

TileOffsets: 324, 0x144.

TileByteCounts: 325, 0x145.

InkSet: 332, 0x14C.

InkNames: 333, 0x14D.

NumberOfInks: 334, 0x14E.

DotRange: 336, 0x150.

TargetPrinter: 337, 0x151.

SampleFormat: 339, 0x153.

SMinSampleValue: 340, 0x154.

SMaxSampleValue: 341, 0x155.

TransferRange: 342, 0x156.

JPEGProc: 512, 0x200.

JPEGInterchangeFormat: 513, 0x201.

JPEGInterchangeFormatLength: 514, 0x202.

JPEGRestartInterval: 515, 0x203.

JPEGLosslessPredictors: 517, 0x205.

JPEGPointTransforms: 518, 0x206.

JPEGQTables: 519, 0x207.

JPEGDCTables: 520, 0x208.

JPEGACTables: 521, 0x209.

YCbCrCoefficients: 529, 0x211.

YCbCrSubSampling: 530, 0x212.

YCbCrPositioning: 531, 0x213.

ReferenceBlackWhite: 532, 0x214.

Die Verarbeitbarkeit in den obigen Listen *nicht* enthaltener Tags ist schwer vorhersehbar, meist aber wohl auf sehr spezifische Software beschränkt.

#### **4.1.4 Anmerkungen**

Neuere Bildverarbeitungssoftware tendiert mehr und mehr dazu einzelne private fields mit Metadaten im weitesten Sinne in TIF Dateien zu schreiben, von denen sich der Hersteller Vorteile verspricht. Da Baseline TIFF Reader angewiesen sind, alle ihnen unbekanntes Felder zu ignorieren, stellt das Vorhandensein derartiger nicht in der Baseline Spezifikation enthaltener Felder kein Risiko dar – *dann, und nur dann, wenn die über diese Nicht-Standardfelder gespeicherte Information keine für den Projektzweck relevanten Informationen enthält.*

## **4.2 Ordnerstruktur**