

Inhalt

1. Executive Summary	3
2. Vorbemerkung	4
3. Digitale Langzeitarchivierung an der ETH Zürich	4
4. Bilddatenformate	4
4.1. Retrodigitalisate und Born Digitals.....	5
4.2. Formatkonvertierungen.....	5
4.3. Tagged Image File Format (TIFF)	6
4.4. RAW-Formate.....	6
4.5. Digital Negative (DNG)	7
4.6. JPG2000.....	7
4.7. Adobe Formate	7
4.8. JPEG, Portable Network Graphics (PNG), Graphics Interchange Format (GIF).....	7
5. Konkretisierung TIFF	8
5.1. Reihenfolge der Tags	8
5.2. Farbtiefe und Auflösung.....	8
5.3. Image File Directories (IFD), Pyramiden und Multipage.....	8
5.4. Speicherung: Streifen und Kacheln	9
5.5. Speicherung der Pixelwerte.....	9
5.6. Komprimierung	9
5.7. Ebenen	10
5.8. Farbprofil.....	10
5.9. TIFF-Tags	10
5.9.1. Verpflichtende Tags	11
5.9.2. Optionale Tags.....	12
6. Metadaten	13
6.1. Technische Metadaten	13
6.1.1. Zwingende technische Metadaten	13
6.1.2. Optionale technische Metadaten.....	13
6.2. Deskriptive- und Rechte-Metadaten	14
6.2.1. IPTC-IIM und XMP-Standard	14
6.2.2. Erfassen von Metadaten in Photoshop und Bridge.....	14
6.2.3. Erfassen bzw. Übernahme von Metadaten in E-Pics.....	15
7. Referenzen	16
8. Anhang.....	17
8.1. Illustration Baseline TIFF-Tags.....	17

1. Executive Summary

Im vorliegenden Dokument werden Empfehlungen für die Langzeitarchivierung von Bilddaten im ETH Data Archive formuliert. Diese zielen darauf ab, dass Bilddaten auch in Zukunft noch angezeigt und nachgenutzt werden können. Die Empfehlungen lassen sich in folgende Bereiche einteilen:

- Formatwahl
- Für die Anzeige durch Viewer notwendige technische Metadaten
- Für die Nachnutzung relevante deskriptive und Rechtemetadaten

Für die Archivierung geeignete Formate sind offen spezifiziert, bieten verlustfreie Komprimierung, sind weit verbreitet und werden damit durch gängige Software gut unterstützt. Es gelten folgende Empfehlungen (vgl. Kapitel 4, Bilddatenformate):

Format	Empfehlung
TIFF	Empfohlen für Digitalisierung und Archivierung.
DNG	Empfohlen für die Archivierung von Rohdaten.
JPG2000	Archivtauglich, wenn verlustfreie Komprimierung genutzt wird. TIFFs werden bevorzugt.
RAW-Formate	Nicht archivtauglich. Empfohlen wird die Konvertierung nach TIFF oder DNG
Adobe Formate	Adobe-spezifische Formate (z.B. Photoshop, InDesign, Illustrator) sind nicht archivtauglich und sollten ins TIFF-Format konvertiert werden.
JPG, PNG, GIF	Weit verbreitet und dadurch zum jetzigen Zeitpunkt nicht gefährdet. Eine Konvertierung ins TIFF-Format wird derzeit nicht empfohlen für die Produktion neuer Dateien zu Archivierungszwecken sollte immer das TIFF-Format verwendet werden.

Für die Langzeitarchivierung hat sich das TIFF-Format als Standard durchgesetzt. Die für die Anzeige notwendigen technischen Metadaten sind im TIFF 6.0 Baseline-Standard definiert. Sie werden durch die Software in Abhängigkeit eingestellter Parameter automatisch in die Datei geschrieben. Es gelten für die Produktion bzw. für das Abspeichern ins TIFF-Format folgende Empfehlungen (siehe Kapitel 5):

- In der Regel ist eine Farbtiefe von 8 Bit und eine Auflösung von 300dpi ausreichend
- Keine Pyramidenspeicherung, keine Multipage-TIFFs, keine Ebenen aus Photoshop
- Speicherung der Bilddaten in Stripes (Streifen), keine Kachelspeicherung
- Keine Komprimierung
- Einbettung des Farbprofils in die TIFF-Datei

Für die langfristige Nutzbarkeit von Bilddatenformaten sind neben den technischen auch deskriptive und Rechtemetadaten relevant (Kapitel 6.2). Diese können im XMP und/oder IPTC-Standard direkt in die TIFF-Datei gespeichert werden.

2. Vorbemerkung

Das vorliegende Dokument stellt einerseits eine für das TIFF-Format verbindliche Richtlinie für die Übernahme von Bilddaten ins Data Archive der ETH Zürich dar und gibt andererseits Empfehlungen für die Langzeitarchivierung von Bilddateien allgemein an die Hand. Die Empfehlungen können sich von den Empfehlungen anderer Langzeitarchive unterscheiden.

In Kapitel 3 wird die Langzeitarchivierung im ETH Data Archive vorgestellt, wobei insbesondere auf die eingesetzte Software und die Speicherinfrastruktur eingegangen wird. Kapitel 4 geht auf gängige Formate ein und gibt konkrete Empfehlungen zur Wahl des geeigneten Dateiformats¹ in Abhängigkeit des Archivierungsziels ab. Das für die Langzeitarchivierung geeignetste und verbreitetste Format ist das TIFF-Format, weswegen diesem das ganze Kapitel 5 gewidmet ist. Erläutert werden für die Langzeitarchivierung wichtige Eigenschaften, welche TIFF-Dateien erfüllen sollten sowie konkrete Hinweise hinsichtlich der wählbaren Parametern beim Abspeichern von TIFF-Dateien. Informationen und Empfehlungen hinsichtlich Metadaten werden in Kapitel 6 behandelt.

3. Digitale Langzeitarchivierung an der ETH Zürich

Das ETH Data Archive ist das Digitale Langzeitarchiv für die ETH Zürich. Es setzt für die Speicherung digitaler Objekte die Archivierungssoftware Rosetta der Firma Ex Libris ein und nutzt ausschliesslich die Speicherinfrastruktur der Informatikdienste der ETH Zürich.

Rosetta identifiziert das Dateiformat und validiert für bestimmte Formate wie das TIFF die Übereinstimmung der jeweiligen Datei mit der Spezifikation des Formats. Zudem werden Prüfsummen generiert, welche es erlauben, die Datenintegrität zu gewährleisten.

Das ETH Data Archive ist in erster Linie ein Archivsystem («dark archive») und keine Publikationsplattform. Die Daten werden in der Regel zunächst in andere Systeme importiert, verwaltet, distribuiert und von dort zusammen mit den Metadaten für den Ingest in Rosetta exportiert.

4. Bilddatenformate

Für die Langzeitarchivierung ist zunächst die Wahl des Formats entscheidend. Die folgenden Kriterien werden allgemein und im ETH Data Archive für die Einschätzung der Aussichten für den Langzeiterhalt von Daten herangezogen. Im Idealfall treffen alle Kriterien zu:

- Weite Verbreitung
- Offene Spezifikation des Formats
- Keine oder verlustfreie Kompression (Algorithmus muss offen dokumentiert sein)

Selbstverständlich ist die Beurteilung dieser Kriterien in regelmässigen Abständen zu überprüfen, da die Dateiformate einem steten Wandel unterworfen sind. Dabei wird der Fokus vor allem auf

¹ Wenn in der Folge nicht ausdrücklich etwas anderes erwähnt wird, ist mit «Format» jeweils das Dateiformat gemeint.

drohende Obsoleszenz gelegt. Obsoleszenz bezeichnet die Konsequenz, die sich aus der technologischen Entwicklung ergibt und dazu führt, dass zum Beispiel Software oder Dateiformate durch neuere Versionen ersetzt und damit obsolet werden. Die veralteten Versionen werden durch Anwender nicht mehr verwendet, womit auch der Support durch Betriebssysteme oder Anwendersoftware immer mehr schwindet womit die Lesbarkeit der gespeicherten Inhalte bedroht ist. Die Frage, die sich in diesem Zusammenhang stellt, ist: Gibt es bedrohte Formate, die durch neuere, effizientere Formate umfassend abgelöst werden? Falls die Frage bejaht wird, muss eine Formatmigration erwogen werden, d.h. die Bildinhalte werden in ein geeignetes, langzeitarchivtaugliches Format übertragen, wobei ein Informationsverlust möglichst verhindert werden muss. Damit wird sichergestellt, dass die Dateien in aktuellen und gängigen Applikationen angezeigt werden können.

Unter den Bilddateiformaten haben sich zum aktuellen Zeitpunkt insbesondere das Tagged Image File Format (TIFF), das Digital Negative von Adobe (DNG) und das JPG-2000 Format als für die Archivierung digitaler Bilddaten geeignet herauskristallisiert².

In den folgenden Kapitel werden die verschiedenen Dateiformate beschrieben und jeweils Empfehlungen formuliert. Doch zunächst wird kurz auf die Unterscheidung zwischen «Retrodigitalisate» und «Born Digitals» sowie auf die Formatmigration allgemein eingegangen.

4.1. Retrodigitalisate und Born Digitals

Im Archivbereich werden digitale Bilder, die als Reproduktion analoger Objekte angefertigt werden, und nativ digitale Bilder unterschieden. Es werden hierfür die Begriffe Retrodigitalisate und Born Digitals verwendet.

4.2. Formatkonvertierungen

Grundsätzlich lässt sich jedes Format durch eine Formatmigration in das für die Langzeitarchivierung besonders günstige TIFF-Format überführen. Bei Ausgangsformaten, die sich besonders schlecht für Langzeitarchivierung eignen, z.B. Rohdaten-Formaten (Kapitel 4.4) oder Adobe eigenen Bearbeitungsformaten (Kapitel 4.7) wird dies in vielen Fällen Sinn machen. Es gibt allerdings Aspekte, die gegen eine Migration sprechen:

- Aufwand: Formatmigrationen sind aufwändig und binden viele Ressourcen
- Verlust von Informationen: Eine Migration kann zum Verlust von wichtigen Eigenschaften führen. Zum Beispiel können Bearbeitungsmöglichkeiten durch die Konvertierung von Adobe-Formaten zu TIFF verloren gehen.
- Veränderungen des Bildinhalts sind nicht ausgeschlossen und müssen oft aufwendig durch Sichtkontrolle ausgeschlossen werden.

² Sämtliche Formatempfehlungen des ETH Data Archive finden sich unter: <https://documentation.library.ethz.ch/display/DD/Archivtaugliche+Dateiformate>, Stand [19.04.2021]

- Verlust der Authentizität: Mit der Migration wird der Datenstrom der Datei verändert und entspricht nicht mehr dem Zustand zum Zeitpunkt der Ablieferung, womit die Authentizität oft nicht mehr gewährleistet werden kann.

Formatkonvertierungen müssen je nach Ausgangslage, was Ressourcen, Datenbestand und Anforderungen an die Archivierung betrifft, sorgfältig abgewogen werden. Spätestens aber, wenn ein Dateiformat von Obsoleszenz bedroht wird, kann eine Migration nicht mehr aufgeschoben werden um die Nutzung der in diesem Format gespeicherten Inhalte gewährleisten zu können.

4.3. Tagged Image File Format (TIFF)

Das TIFF-Format ist sehr gut dokumentiert und wird verbreitet eingesetzt. In Langzeitarchiven hat sich TIFF 6.0 (Baseline) längst als Standard für die Langzeitarchivierung durchgesetzt. Mit dem Baseline-Standard wurde ein Minimalstandard mit den für die Anzeige zwingenden technischen Metadaten geschaffen (siehe auch Kapitel 4.7.1), welche von jeder Applikation interpretiert werden müssen. Sämtliche TIFF Baseline Tags inkl. der jeweils erlaubten Werte finden sich in der TIFF 6.0 Spezifikation unter Part 1: Baseline TIFF³. Eine sehr übersichtliche Zusammenstellung der Baseline Tags findet sich ausserdem bei Aware Systems⁴. Die Empfehlungen des ETH Data Archives schliessen den Baseline-Standard vollständig ein.

Empfehlung: Für Retrodigitalisate sollte in jedem Fall das TIFF-Format gewählt werden. Bei Born-Digital-TIFFs sollten die in Kapitel 5 genannten Eigenschaften überprüft und allenfalls analog den Empfehlungen neu abgespeichert werden. Born-Digital-Dateien anderer Formate muss eine Migration ins TIFF-Format geprüft werden.

4.4. RAW-Formate

Als Rohdatenformat bezeichnet man das jeweilige herstellerspezifische Dateiformat bei Digitalkameras, mit denen die Kamera die reinen Sensordaten nach der Aufnahme auf das Speichermedium schreibt. In digitalen Kameras kann eingestellt werden, ob Aufnahmen in RAW oder als JPG bzw. in anderen Formaten gespeichert werden sollen. Vorteil von Rohdaten sind die weitreichenden Möglichkeiten und Reserven in der Bildbearbeitung. Rohdaten können aber nur mit einem RAW-Konverter angezeigt und bearbeitet werden, weswegen sie in der Regel nach Abschluss der Bearbeitung in ein gängiges Bilddateiformat exportiert werden. Rohdaten-Formate, manchmal auch als «digitale Negative» bezeichnet, sind proprietär und unterscheiden sich von Hersteller zu Hersteller, manchmal von Modell zu Modell und im Extremfall selbst zwischen verschiedenen Versionen eines Kameramodells. Ein Standardformat gibt es bis heute nicht. Adobe versucht immerhin mit dem DNG-Format (Kapitel 4.5) einen Standard zu etablieren, bislang allerdings ohne durchschlagenden Erfolg. Rohdatenformate eignen sich nicht für die Langzeitarchivierung und sollten nach erfolgter Bearbeitung ins TIFF-Format, oder, wenn die Bilddateien weiterhin bearbeitet werden müssen, ins DNG-Format konvertiert werden.

³ TIFF Revision 6.0, abgerufen unter: <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf>, Stand [19.04.2021]

⁴ TIFF Tag Reference, Baseline TIFF Tags, <https://www.awaresystems.be/imaging/tiff/tifftags/baseline.html>, Stand [19.04.2021]

Empfehlung: RAW-Formate wie CR2, NEV, RAV ARW usw. sollten für die Archivierung ins TIFF- oder, wenn die Datei für die Nachbearbeitung vorgehalten wird, ins DNG Format konvertiert werden! Für die Konvertierung ins TIFF-Format sind die Empfehlungen in Kapitel 5 zu beachten.

4.5. Digital Negative (DNG)

Das «Digital Negative», kurz DNG, wurde von Adobe entwickelt, um ein Standardformat für Rohdaten (vgl. Kapitel 4.4) zur Verfügung zu stellen. Leider konnte sich das DNG bislang nicht durchsetzen. Gleichwohl eignet es sich aufgrund der offen verfügbaren Formatspezifikation grundsätzlich für die Langzeitarchivierung.

Empfehlung: Grundsätzlich ist das DNG archivtauglich und kann für Rohdaten, die weiter bearbeitet werden sollen, eingesetzt werden. Für reine Archivierung wird das TIFF-Format dem vorgezogen.

4.6. JPG2000

Das JPG2000-Format eignet sich für die Langzeitarchivierung, wenn es mit verlustfreier Komprimierung verwendet wird. JPG2000 ist im Vergleich zum TIFF-Format aber weniger verbreitet, weswegen das TIFF-Format klar bevorzugt wird.

Empfehlung: Bilddaten im Format JPG2000 mit verlustfreier Komprimierung sind grundsätzlich archivtauglich, das TIFF-Format sollte bevorzugt werden.

4.7. Adobe Formate

Die proprietären Adobe-Formate, welche in den Programmen Photoshop (.psd), InDesign (.indd), Illustrator (.ai) usw. eingesetzt werden, sind für die Langzeitarchivierung nicht geeignet und sollten entweder ins TIFF-Format oder, wenn die weitere Bearbeitung vorgesehen ist, ins DNG-Format konvertiert werden.

Wenn es sich um Daten handelt, an denen weiterhin gearbeitet wird kommt eine Konvertierung natürlich nicht in Frage, da sämtliche Bearbeitungsmöglichkeiten hinfällig würden. Nach Abschluss der Arbeiten sollten die Ergebnisse für die Archivierung als TIFF- oder DNG gespeichert werden.

Empfehlung: Proprietäre Formate von Adobe Produkten (Photoshop, InDesign, Illustrator) sollten zwecks Langzeitarchivierung ins TIFF- oder DNG Format konvertiert werden!

4.8. JPEG, Portable Network Graphics (PNG), Graphics Interchange Format (GIF)

JPEG, PNG und GIF sind verlustbehaftet komprimierte Dateiformate. Aufgrund ihrer grossen Verbreitung werden sie durch gängige Betrachtungstools so gut unterstützt, dass ihre Nutzbarkeit aktuell trotz eingesetzter Komprimierung nicht durch Obsoleszenz gefährdet ist. Trotzdem werden sie für eine längere Aufbewahrung nur als bedingt geeignet eingestuft.

Empfehlung: JPG-, PNG- und GIF-Dateien sind bedingt archivtauglich. Aktuell wird keine Konvertierung ins TIFF-Format empfohlen.

5. Konkretisierung TIFF

In diesem Kapitel werden für die Archivierung wichtige Eigenschaften von TIFF-Dateien erläutert und entsprechende Empfehlungen ausgesprochen.

5.1. Reihenfolge der Tags

Die in der TIFF-Datei eingebetteten Metadaten-Tags müssen gemäss TIFF-Baseline-Spezifikation in aufsteigender Reihenfolge vorliegen. In der Regel speichern gängige Tools die Tags entsprechend dieser Vorgabe korrekt ab und es besteht kein Handlungsbedarf. Falls es trotzdem Unregelmässigkeiten gibt, kann die Reihenfolge der Tags mittels erneutem Abspeichern mit Photoshop oder mittels Neuschreiben der Tags mit Exiftool behoben werden.

5.2. Farbtiefe und Auflösung

Das ETH Data Archive empfiehlt eine Farbtiefe von 8 Bit und eine Auflösung von 300dpi⁵. Je nach Art, Grösse, Zustand des Originals, Details im Bildinhalt und Zweck der Digitalisierung sind Abweichungen nach oben sinnvoll (16 Bit, 600dpi⁶).

5.3. Image File Directories (IFD), Pyramiden und Multipage

Ein Bild (mit bestimmter Auflösung) wird im TIFF-Format in einem «Image File Directory» (IFD) gespeichert. In jedem Fall wird das «Hauptbild» mit voller Auflösung in einem IFD abgespeichert. Daneben ist es möglich, weitere IFDs vorzuhalten, zum Beispiel:

- mit *einem* Vorschaubild mit niedriger Auflösung
- mit *mehreren* niedriger aufgelösten Bildern (sogenannte Bildpyramide)
- weiteren Bildern voller Auflösung (Multipage TIFFs)

Da ein TIFF-Baseline-konformer Reader lediglich das erste IFD auslesen muss, besteht einerseits das Risiko, dass weitere im File enthaltene relevante Inhalte nicht angezeigt werden, andererseits, dass aufgrund der Reihenfolge der IFDs nicht zwingend das Masterbild in voller Auflösung sondern lediglich ein Vorschaubild angezeigt wird. Auf TIFF-Dateien mit eingebetteten Vorschaubildern, Pyramiden- oder Multipage TIFFs⁷ sollte deswegen verzichtet werden.

⁵ Die Angabe der Auflösung in dpi bezieht sich auf die Grösse des gedruckten oder gescannten Dokumentes. Sie muss von den eigentlichen Bilddimensionen (Pixelbreite und -höhe) unterschieden werden. Die Druckgrösse (hier 300dpi) gibt die Anzahl Pixel per inch des gescannten oder gedruckten Dokumentes an. Für die Wiedergabe am Bildschirm muss das Dokument auf dem Bildschirm auf die Originalgrösse gezoomt werden, damit die Auflösung in dpi auf dem Bildschirm korrekt angezeigt wird.

⁶ Die Datenmenge bei 16 Bit Farbtiefe kann entsprechend hohe Anforderungen an die Infrastruktur für die Nutzung dieser Dateien stellen.

⁷ Multipage-TIFFs sind gemäss KOST weit verbreitet und grundsätzlich nicht ungeeignet für die Archivierung, im Data Archive werden sie aber aus den genannten Gründen grundsätzlich abgelehnt.

TIFF-Dateien dürfen nur genau ein IFD mit dem Masterbild enthalten. In Photoshop muss bei der Speicherung einer TIFF-Datei die Checkbox für die «Bildpyramide» deaktiviert sein.

5.4. Speicherung: Streifen und Kacheln

TIFF-Dateien sollten in Streifen (Stripes) abgespeichert werden, weil die Speicherung in Kacheln («Tiles») nicht Baseline-konform ist und aufgrund geringer Verbreitung ein Risiko darstellt.

Hintergrundinformation: Bei Speicherung von Streifen werden die Tags 273, 278 und 279 geschrieben bzw. müssen zwingend vorhanden sein, wohingegen bei der Speicherung mit Kacheln die Tags 322 bis 325 geschrieben werden.

TIFF-Dateien immer mit Streifen (Strips) abspeichern. Die Speicherung als Kacheln oder eine Mischform von beidem muss vermieden werden.

5.5. Speicherung der Pixelwerte

Pixelwerte können auf unterschiedliche Weise gespeichert werden. Bei RGB Bildern kann jeder Farbwert nacheinander (RGB-RGB-RGB...) oder jeder Farbkanal separat gespeichert werden. Erstere Speicherungsart wird in Photoshop Interleaved bezeichnet und muss in jedem Fall bevorzugt werden, da letztere viel weniger verbreitet ist und die korrekte Interpretation durch Viewer nicht gewährleistet werden kann, da diese nicht im Baseline-Standard spezifiziert ist.

Wenn TIFF-Dateien in Photoshop gespeichert werden, muss bei «Pixel Order» zwingend «interleaved» angewählt werden.

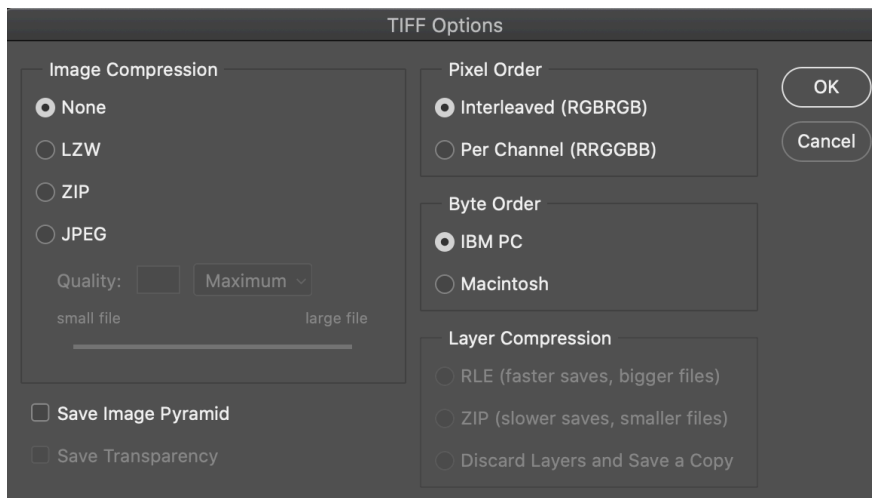
Hintergrundinformation: Eine korrekt gespeicherte TIFF-Datei enthält im Tag 284 (PlanarConfiguration) den Wert «Chunky». Eine nicht Baseline-konform gespeicherte TIFF-Datei dagegen enthält in Tag 284 (PlanarConfiguration) den Wert «Planar» und weiter in den Tags 273 und 279 (StripOffsets bzw. StripByteCounts) drei Werte statt nur einem.

5.6. Komprimierung

Akzeptiert werden TIFF-Dateien mit der verlustfreien Komprimierung mittels des Lempel-Ziv-Welch-Algorithmus (LZW)⁸. Für die Langzeitarchivierung wird allerdings empfohlen, gänzlich auf eine Komprimierung zu verzichten und stets unkomprimiert abzuspeichern (Tag 259, Compression, enthält den Wert 1). Verlustfrei komprimierte TIFFs können durch erneutes Abspeichern ohne Qualitätsverlust in unkomprimierte TIFFs umgewandelt werden. Bei verlustbehafteter Komprimierung kann dadurch der Qualitätsverlust nicht mehr rückgängig gemacht werden.

Wenn TIFF-Dateien in Photoshop gespeichert werden, muss bei den Kompressionsoptionen «none» angewählt werden. In Ausnahmefällen, wenn eine Komprimierung aufgrund sehr hoher Speichermenge notwendig ist, ist LZW auszuwählen.

⁸ Eine Übersicht über die Komprimierungsverfahren findet sich bei der Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): <https://kost-ceco.ch/cms/tiff-tagged-image-file-format-1.html>, Stand [19.04.2021]

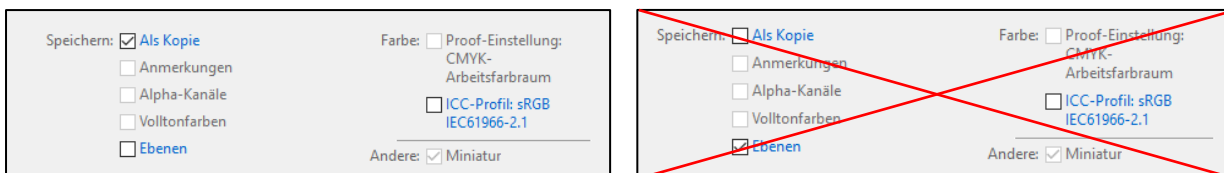


Screenshot: Photoshop Parameter TIFF-Speicherung

5.7. Ebenen

Auf den Einsatz von Ebenen muss verzichtet werden, da der korrekte Umgang und damit die korrekte Darstellung des Bildinhalts durch verschiedene Softwareprodukte nicht gewährleistet werden kann. Bei der Nachbearbeitung ist Vorsicht geboten, denn Photoshop speichert evtl. vorhandene Ebenen unter Umständen standardmässig ab. Die entsprechenden Checkboxes sollte daher unbedingt beachtet werden.

- Checkbox «Als Kopie» muss aktiviert sein
- Checkbox «Ebenen» darf nicht aktiviert sein



Screenshots: Photoshop Optionen für TIFF-Speicherung

5.8. Farbprofil

Falls ein Farbprofil, auch ICC-Profil (ICC: International Color Consortium) vorhanden ist, sollte dieses in den entsprechenden Tag 34675 geschrieben werden⁹.

5.9. TIFF-Tags

Bei den verpflichtenden Tags (Kapitel 5.9.1) handelt es sich weitgehend um Tags, welche Teil des TIFF-Baseline-Standards darstellen und für die korrekte Anzeige durch TIFF-fähige Software

⁹ Auch wenn mit einem ICC-Profil die Farbgenauigkeit gewährleistet werden kann, werden nicht alle Tools das ICC-Profil aus: IrfanView, Windows Media Player, Paint, Paint 3D, XnView zum Beispiel stellen Farbtöne unter Umständen verfälscht dar.

zwingend notwendig sind. Die im TIFF-Baseline-Standard definierten Tags müssen von jedem TIFF-Reader gelesen und interpretiert werden können. Die Orientierung an diesen Baseline-TIFF-Tags stellt die zwingende Grundlage für den erfolgreichen Langzeiterhalt von Bilddaten dar. Sämtliche TIFF-Baseline-Tags inkl. der jeweils erlaubten Werte finden sich in der TIFF 6.0 Spezifikation unter Part 1: Baseline TIFF¹⁰. Eine sehr übersichtliche Zusammenstellung der Baseline Tags stellt Joris Van Damme auf seiner Website zur Verfügung¹¹. Eine weitere wichtige Grundlage für die Definition von zwingenden und optionalen Tags sind die Richtlinien der SLUB Dresden (2019)¹².

Die in den verpflichteten Tags gespeicherten technischen Metadaten werden von Hard- und Software automatisch geschrieben und sind teilweise von den bei der Digitalisierung und Bildbearbeitung verwendeten Parametern abhängig.

Die optionalen Tags sind für die Anzeige des Bildes entweder nicht notwendig oder müssen nicht vorhanden sein, weil TIFF-fähige Programme Baseline-konforme Default-Werte annehmen. Diese aufgeführten optionalen Tags können bei Bedarf verwendet werden, müssen es aber nicht. Wenn sie verwendet werden, sind die zulässigen Werte zu beachten.

Tags, die hier nicht aufgeführt werden, sind grundsätzlich nicht erlaubt. Wenn die Notwendigkeit für solche Tags besteht, muss die Verwendung vorgängig mit dem ETH Data Archive abgesprochen werden.

5.9.1. Verpflichtende Tags

Die hier aufgeführten Tags beschreiben technische Parameter, die für die Anzeige des digitalisierten Bildes vorhanden sein müssen. Da die Anzeigefähigkeit die zentrale, langfristig zu erhaltende Funktion eines Bildes ist, müssen diese Tags auch für die Gewährleistung des langfristigen Erhalts des Bildes zwingend vorhanden sein. Fehlt einer dieser Tags oder ist der Informationsgehalt zum Beispiel bezüglich des erforderlichen Datentyps nicht Baseline-konform geschrieben, kann die korrekte Anzeige des Inhalts nicht mehr gewährleistet werden. Da die Baseline-Tags sehr gut dokumentiert und genau spezifiziert sind, erkennen gängige Validatoren diese Fehler zuverlässig.

Tabelle: Verpflichtende Tags

Dez.	Hex.	Bezeichnung	Erlaubte Werte	Erläuterung	Ref.
256	0x0100	ImageWidth	Wert zwischen 1 und 4.294.967.295	Bildbreite in Pixel	Link
257	0x0101	Image Length	Wert zwischen 1 und 4.294.967.295	Bildhöhe in Pixel	Link
258	0x0102	BitsPerSample	RGB-Bilder: 8, 16, Graustufen: 1, 4, 8	Bittiefe (Anzahl darstellbarer Farbtöne)	Link
259	0x0103	Compression	1 (Defaultwert, keine Kompr.)	Komprimierung	Link

¹⁰ TIFF Revision 6.0, abgerufen unter: <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf>, Stand [19.04.2021]

¹¹ TIFF Tag Reference, Baseline TIFF Tags, <https://www.awaresystems.be/imaging/tiff/tifftags/baseline.html>, Stand [19.04.2021]

¹² https://www.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Handreichung_TIFF_v1.3.5.pdf, Stand [28.04.2021]

262	0x0106	PhotometricInterpretation	0, 1, 2	Farbraum (2 = RGB)	Link
273	0x0111	StripOffsets	Wert zwischen 1 und 4.294.967.295	Offsets der Pixelzeilen	Link
277	0x0115	SamplesPerPixel	RGB: 3, Graustufen, Bitonal: 1 (Defaultwert)	Anzahl Komponenten pro Pixel	Link
278	0x0116	RowsPerStrip	Wert zwischen 1 und 4.294.967.295	Anzahl Pixelzeilen	Link
279	0x0117	StripByteCounts	Wert zwischen 1 und 4.294.967.295	Gesamtzahl Pixel; ergibt sich aus den Tags 256, 257 und 277	Link
282	0x011a	XResolution	>= 300	Horizontale Auflösung	Link
283	0x011b	YResolution	>= 300	Vertikale Auflösung	Link
34675		InterColourProfile		ICC-Profil	Link

Legende:

Werte abhängig von Einstellungen beim Speichern

5.9.2. Optionale Tags

Die optionalen Tags sind entweder für die Anzeige des Bildes nicht notwendig oder müssen nicht vorhanden sein, weil TIFF-fähige Programme bei Abwesenheit der Tags Baseline-konforme Standardwerte annehmen. Diese Tags können bei Bedarf verwendet werden, müssen es aber nicht.

Die untenstehend definierten optionalen Tag entsprechen weitgehend den von der SLUB Dresden definierten. Im Unterschied dazu werden unerwünschte bzw. verbotene Tags hier nicht aufgelistet.

Es gilt der Grundsatz, dass alle hier nicht aufgeführten Tags einer Abklärung durch das ETH Data Archive bedürfen.

Tabelle: Optionale Tags

Dez.	Hex.	Bezeichnung	Erlaubte Werte	Erläuterung	Ref.
254	0x00FE	NewSubFileType	0	Art der Bilddaten	Link
266	0x010A	FillOrder	1 (= Defaultwert)	Art der Speicherung	Link
269	0x010D	DocumentName	String in ASCII	Titel	Link
270	0x010E	ImageDescription	String in ASCII	Beschreibung	Link
271	0x010F	Make	String in ASCII	Gerätehersteller	Link
272	0x0110	Model	String in ASCII	Gerätemodel	Link
274	0x0112	Orientation	1 (= Defaultwert)	Bildausrichtung, wenn vorhanden, nur Wert 1 erlaubt	Link
280	0x0118	MinSampleValue	Keine Einschränkungen	Statistische Angabe	Link

281	0x0119	MaxSampleValue	Keine Einschränkungen	Statistische Angabe	Link
284	0x011C	PlanarConfiguration	1 (Defaultwert)	Art, wie Pixelwerte gespeichert werden, wenn vorhanden, nur Wert 1 erlaubt	Link
290	0x0122	GrayResponseUnit		Einheit für die GrayResponseCurve (Tag 291)	Link
291	0x0123	GrayResponseCurve		Optische Dichte für jeden Pixelwert von Graustufenbildern	Link
296	0x0128	ResolutionUnit	2 (Inch; Defaultwert)	Einheit Auflösung	Link
297	0x0129	PageNumber	Optional	Seitenangabe	Link
305	0x0131	Software	String in ASCII	Erstellungs-Software	Link
306	0x0132	DateTime	Format: YYYY:MM:DD HH:MM:SS	Erstellungsdatum und -Uhrzeit	Link
318	0x013E	WhitePoint		Weisspunkt	Link
319	0x013F	PrimaryChromaticities		Farbwerte der Primärfarben	Link
339	0x0153	SampleFormat	1 (Defaultwert)	1 = unsigned integer	Link
33432	0x8298	Copyright	String in ASCII	Copyrightvermerk	Link

6. Metadaten

Metadaten sind für die Langzeitarchivierung von entscheidender Bedeutung. Technische Metadaten befähigen ein Tool, den Bildinhalt korrekt darzustellen, deskriptive Metadaten machen den Bildinhalt und evtl. den Kontext des Bildes verständlich. Rechte-Metadaten geben Auskunft über die Nutzungsmöglichkeiten, sie sind für die Nachnutzung entscheidend. Die für die Darstellung des Bildinhalts notwendigen technischen Metadaten werden von der Software automatisch geschrieben, während die deskriptiven- und Rechtemetadaten manuell erfasst werden müssen.

6.1. Technische Metadaten

6.1.1. Zwingende technische Metadaten

Sämtliche technischen Metadaten, also Angaben, welche eine Software für die Anzeige des Bildinhalts zwingend benötigt, werden automatisch durch die erzeugenden Geräte und Software-Anwendungen anhand der eingestellten Parameter in die dafür vorgesehenen Tags geschrieben (vgl. Kapitel 4.7.1).

6.1.2. Optionale technische Metadaten

Nach der Bildbearbeitung werden durch die benutzte Software technische Metadaten in die resultierende Bilddatei abgespeichert. Wenn zum Beispiel eine TIFF-Datei mit Photoshop bearbeitet und/oder abgespeichert wird, schreibt Photoshop eigene Metadaten als XMP in die Datei¹³. Diese für Photoshop spezifischen Metadaten stellen für die Langzeitarchivierung weder ein Risiko noch

¹³Siehe auch Kapitel 6.2.1. Photoshop-Tags sind hier dokumentiert: <https://exiftool.org/TagNames/XMP.html#photoshop>, Stand [19.04.2021]

einen Nutzen dar, führen aber zu einer tendenziell unübersichtlicheren Datei. Es wird empfohlen, die Photoshop-Metadaten mithilfe des Validierungstools «ExifTool»¹⁴ zu löschen, um optimale Voraussetzungen für die Langzeitarchivierung seiner digitalen Bestände zu schaffen.

6.2. Deskriptive- und Rechte-Metadaten

Deskriptive- und Rechte-Metadaten werden im Unterschied zu den technischen Metadaten nicht automatisch durch die für die Erstellung und Bearbeitung verwendeten Softwares geschrieben. Sie müssen intellektuell erfasst werden. Dabei gibt es mehrere Metadatenstandards, die genutzt werden können und eine Vielzahl unterschiedlicher Tools, die benutzt werden und die natürlich nicht in ihrer Gesamtheit aufgezählt werden können. Im folgenden werden die gängigen Metadatenstandards IPTC-IIM und XMP kurz erläutert.

6.2.1. IPTC-IIM und XMP-Standard

Für die Speicherung deskriptiver Metadaten hat sich seit 1991 das «Information Interchange Modell», der IPTC-IIM-Standard durchgesetzt, der sich stetig weiterentwickelt hat und heute als «IPTC Photo Metadata Standard» in der Version 2019.1 vorliegt¹⁵. Seit 2002 besteht ausserdem der von Adobe entwickelte Standard «Extensible Metadata Platform» (XMP)¹⁶ welcher als ISO-Standard 16684-1 Teil des «IPTC Photo Metadata Standard» wurde. Beide Standards bestehen nebeneinander und werden zumindest von den Adobe Softwareprodukten Photoshop und Bridge gleichzeitig für die Speicherung von Metadaten genutzt, das bedeutet, dass für dieselbe Information sowohl das IPTC- als auch das entsprechende XMP-Feld verwendet wird (vgl. Kapitel 5.2.2).

Es ist wichtig zu wissen, dass die Standards unterschiedlich konzipiert und spezifiziert sind. Relevant sind insbesondere die beim IPTC-IIM-Standard vorhandenen Limitationen was die erlaubte Anzahl Zeichen in den Metadatenfeldern betrifft¹⁷:

Feld		Zeichenlimit (nur IPTC-Core)
Description	Beschreibung	2000
Keywords	Schlagwörter	64 (Pro Schlagwort)
Credit Line	Fotograf	32
Copyright	Copyright	128
Titel/Object Name	Titel	64

6.2.2. Erfassen von Metadaten in Photoshop und Bridge

Sowohl in Adobe Photoshop als auch Bridge wird das Metadateneingabeformular über das *Menu File > Fileinfo* aufgerufen. Die Metadaten werden wie bereits erwähnt nach IPTC und nach XMP abgespeichert.

¹⁴ <https://exiftool.org/> Stand [19.04.2021]

¹⁵ <https://iptc.org/standards/>, Stand [19.04.2021]

¹⁶ <https://www.adobe.com/products/xmp.html>, Stand [19.04.2021]

¹⁷ Angaben übernommen von: https://www.controlledvocabulary.com/imagetdatabases/iptc_naa.html, Stand [19.04.2021]

Damit Software, die lediglich Metadaten im IPTC-Core-Standard unterstützt, sinnvoll genutzt werden kann, sollten bei der Erfassung von Metadaten die geltenden Limiten für IPTC, wie sie in Kapitel 6.2.1 aufgeführt wurden, beachtet werden.

6.2.3. Erfassen bzw. Übernahme von Metadaten in E-Pics

Wird ein DAM wie Cumulus genutzt, muss vorgängig geklärt werden, welche Metadatenfelder auf welche Weise importiert werden. Zu klären ist also, ob die Metadaten vor dem Import in einer beliebigen anderen Anwendung (zum Beispiel Adobe Bridge) verwaltet und mit Metadaten versehen werden oder ob die Bildbestände direkt im DAM organisiert werden. Es gibt grundsätzlich die folgenden Möglichkeiten, Metadaten in E-Pics zu erfassen¹⁸:

- Automatische Übernahme der sich in der Bilddatei befindlichen Metadaten anhand eines vorgängig definierten Mappings beim Import der Metadaten
- Mittels einer CSV- oder TXT-Datei, die manuell angelegt oder aus Excel bzw. einem anderen Tool exportiert wurde (z.B. bei Migrationen).
- Durch manuelle Eingabe während des Imports oder vor dem Import bei der Erstellung eines reinen Metadaten-Satzes

Achtung: E-Pics unterstützt ausschliesslich den IPTC-Core-Standard und übernimmt entsprechend jeweils die maximal erlaubte Anzahl Zeichen je Metadatenfeld. Beachten Sie diesen Umstand, wenn Sie Metadaten mit Tools erfassen, welche unter Umständen keine Warnung bei Überschreitung der maximal zulässigen Zeichenzahl ausgeben (zum Beispiel Adobe Photoshop oder Bridge, vgl. Kapitel 5.3.1).

¹⁸ Weitere Informationen zu E-Pics unter: <https://documentation.library.ethz.ch/display/EPICS/Inhaltsverzeichnis>, Stand [19.04.2021]

7. Referenzen

Aldus Corporation, 1992: TIFF Revision 6.0 Final – June 3, 1992, abgerufen unter: <https://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf> [28.04.2021]

Art1Pirat Blog, 2013: baseline TIFF, abgerufen unter <http://art1pirat.blogspot.com/2013/07/baseline-tiff.html> [28.04.2021]

AWare Systems: TIFF Tag Reference Baseline TIFF Tags, abgerufen unter: <https://www.aware-systems.be/imaging/tiff/tiff-tags/baseline.html> [28.04.2021]

Library of Congress: Sustainability of Digital Formats: Planning for Library of Congress Collection – Still Images >> Tags for TIFF, DNG, and Related Specifications, abgerufen unter: https://www.loc.gov/preservation/digital/formats/content/tiff_tags.shtml [28.04.2021]

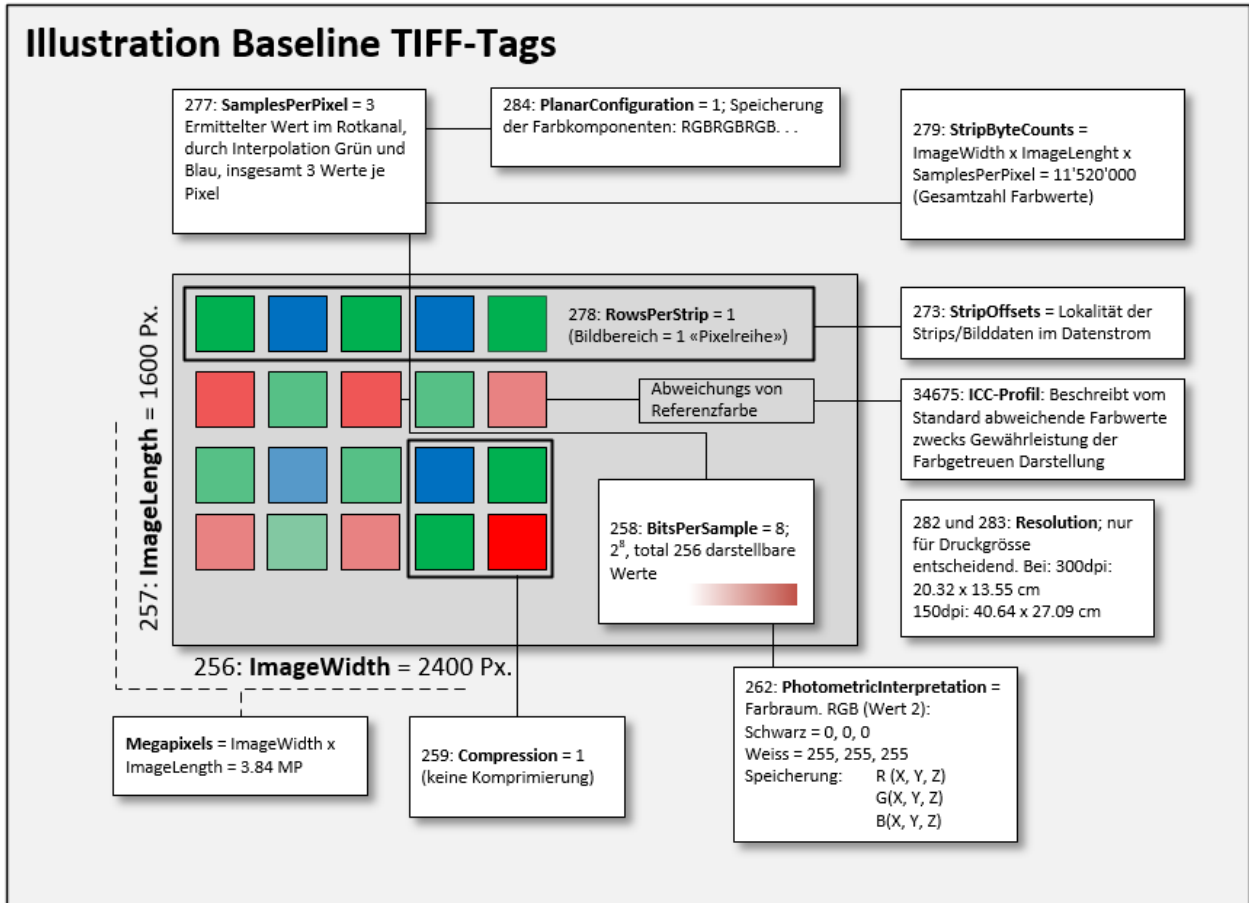
Memoriav, 2017: Memoriav Empfehlungen 2017 FOTO – Die Erhaltung von Fotografien, abgerufen unter: <https://memoriav.ch/wp-content/uploads/2017/11/Memoriav-empfehlungen-Foto-2017.pdf> [28.04.2021]

SLUB Dresden, 2019: Handreichung TIFF, Version 1.3.5, 2019-10-10, abgerufen unter: https://www.slub-dresden.de/fileadmin/groups/slubsite/slubarchiv/SLUBArchiv_Handreichung_TIFF_v1.3.5.pdf [28.04.2021]

Phil Harvey, 2020: XMP Tags, abgerufen unter: <https://exiftool.org/TagNames/XMP.html> [28.04.2021]

8. Anhang

8.1. Illustration Baseline TIFF-Tags



Quelle: ETH Data Archive, 2020