

# TIFF – typische Problemfälle erkennen und beheben

Andreas Romeyke

Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden

2016-03-02

# Überblick

Vor- und Nachteile TIFF

Problemfälle

Werkzeuge

Zusatz: checkit\_tiff

# Überblick

Vor- und Nachteile TIFF

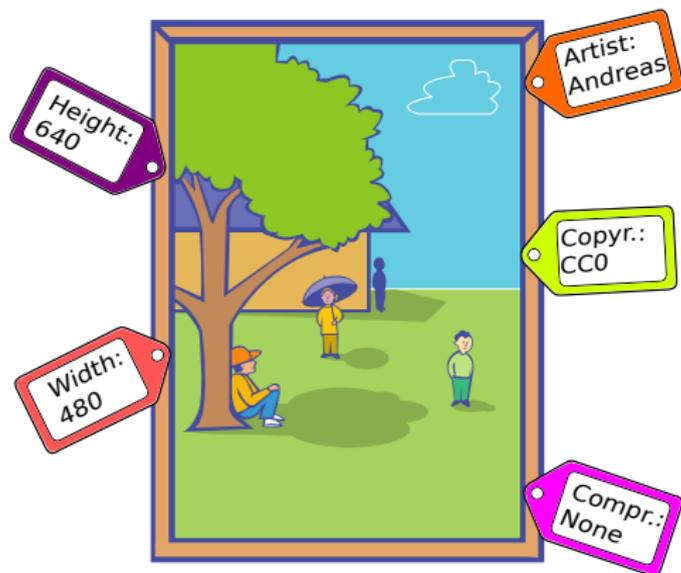
Problemfälle

Werkzeuge

Zusatz: checkit\_tiff

# Was ist TIFF?

Eigenes Werk, verwendet „Etiquette tag icon label“ von „Improulx“ und „kids framed scene“ von „ryanlerch“, Quelle: [openc11part.org](http://openc11part.org)



- ▶ basiert auf
  - ▶ Image File Directories (IFD)
    - ≡ Verzeichnissen
  - ▶ und Tags (Etiketten/Eigenschaften)
- ▶ ...dadurch sehr flexibles Dateiformat

# Spezifikationen TIFF

## TIFF 6.0

- ▶ `ftp://ftp.adobe.com/pub/adobe/devrelations/devtechnotes/pdffiles/tiff6.pdf`
- ▶ bzw. `http://partners.adobe.com/public/developer/tiff/index.html`
- ▶ ergänzt um TIFF Technical Notes, zB. zu 'deflate'-Kompression, ua.

## Sonstiges TIFF

- ▶ bigtiff (<http://bigtiff.org/>)
- ▶ geotiff (<http://www.remotesensing.org/geotiff/spec/geotiffhome.html>)
- ▶ digital negative (DNG) (<https://helpx.adobe.com/photoshop/digital-negative.html>)
- ▶ TIFF/A (in Planung) (<http://www.ti-a.org>)

## Besonderheiten

- ▶ genau ein IFD
- ▶ keine Vorschau-Bilder
- ▶ 12 Pflichttags:  
ImageWidth (256), ImageHeight (257), BitsPerSample (258),  
Compression (259), PhotometricInterpretation (262),  
StripOffsets (273), SamplesPerPixel (277), RowsPerStrip  
(278), StripByteCounts (279), XResolution (282),  
YResolution (283) und ResolutionUnit (296)
- ▶ **private** Tags müssen **nicht** ausgewertet werden
- ▶ keine Kacheln, nur Slices
- ▶ strengere Vorgaben für Tags und Default-Werte

## Besonderheiten

- ▶ genau ein IFD
- ▶ keine Vorschau-Bilder
- ▶ 12 Pflichttags:  
ImageWidth (256), ImageHeight (257), BitsPerSample (258),  
Compression (259), PhotometricInterpretation (262),  
StripOffsets (273), SamplesPerPixel (277), RowsPerStrip  
(278), StripByteCounts (279), XResolution (282),  
YResolution (283) und ResolutionUnit (296)
- ▶ **private** Tags müssen **nicht** ausgewertet werden
- ▶ keine Kacheln, nur Slices
- ▶ strengere Vorgaben für Tags und Default-Werte

## Besonderheiten

- ▶ genau ein IFD
- ▶ keine Vorschau-Bilder
- ▶ 12 Pflichttags:  
ImageWidth (256), ImageHeight (257), BitsPerSample (258),  
Compression (259), PhotometricInterpretation (262),  
StripOffsets (273), SamplesPerPixel (277), RowsPerStrip  
(278), StripByteCounts (279), XResolution (282),  
YResolution (283) und ResolutionUnit (296)
- ▶ **private** Tags müssen **nicht** ausgewertet werden
- ▶ keine Kacheln, nur Slices
- ▶ strengere Vorgaben für Tags und Default-Werte

# Vorteile

- ▶ robustes Dateiformat (sh. auch Offset, tagorder)
- ▶ sauber spezifiziert
- ▶ überschaubare Spezifikation
- ▶ bewährt
- ▶ leicht (und definiert) erweiterbar
- ▶ Definition Kerndaten (baseline TIFF)

# Nachteile

- ▶ nicht streambar
- ▶ unzählige Tags in den Erweiterungen
- ▶ viele fehlerhafte Implementierungen noch in freier Wildbahn
- ▶ Größenbeschränkung auf  $2^{32}$  Bytes (4GiB)
- ▶ bisher kein dediziertes LZA Subset
- ▶ kein formaler Standard

# Überblick

Vor- und Nachteile TIFF

Problemfälle

Werkzeuge

Zusatz: checkit\_tiff

# IFD not in ascending order

## Image File Directory

- ▶ **Alle Tags müssen aufsteigend sortiert sein** (sh. TIFF-Spezifikation, S. 15)
- ▶ falsche Sortierung wäre Hinweis auf Bitfehler, hier:  
**Softwarefehler**
- ▶ vorgekommen in Microfilmscanner (Mekel) von Zeutschel OM1600-HX-4775, Software Quantumscan 1.02.19

## Mögliche Reparatur

- ▶ mit `fixit_tiff`:  
`fixit_tiff -t -i broken_infile.tif -o possibly_repaired_outfile.tif`

# IFD not in ascending order

## Image File Directory

- ▶ Alle Tags müssen aufsteigend sortiert sein (sh. TIFF-Spezifikation, S. 15)
- ▶ falsche Sortierung wäre Hinweis auf Bitfehler, hier:  
**Softwarefehler**
- ▶ vorgekommen in Microfilmscanner (Mekel) von Zeutschel OM1600-HX-4775, Software Quantumscan 1.02.19

## Mögliche Reparatur

- ▶ mit `fixit_tiff`:  
`fixit_tiff -t -i broken_infile.tif -o possibly_repaired_outfile.tif`

# Farbpalette trotz Graustufen

## Tag ColorMap (320) vorhanden

- ▶ Tag PhotometricInterpretation (262) mit Werten 0 bzw. 1 (“min is black” oder “min is white”) signalisiert Graustufen bzw. Monochrome.
- ▶ Vorhandensein von **Tag ColorMap (320) ist dann logischer Widerspruch**, da Vorhandensein von Farbpalette signalisiert wird (sh. TIFF-Spezifikation, S. 37 bzw. S. 23)
- ▶ Hinweis auf **Fehler in der Implementierung**
- ▶ vorgekommen in IrfanView 4.40

## Mögliche Reparatur

- ▶ zZ. unbekannt

# Farbpalette trotz Graustufen

## Tag ColorMap (320) vorhanden

- ▶ Tag PhotometricInterpretation (262) mit Werten 0 bzw. 1 (“min is black” oder “min is white”) signalisiert Graustufen bzw. Monochrome.
- ▶ Vorhandensein von **Tag ColorMap (320) ist dann logischer Widerspruch**, da Vorhandensein von Farbpalette signalisiert wird (sh. TIFF-Spezifikation, S. 37 bzw. S. 23)
- ▶ Hinweis auf **Fehler in der Implementierung**
- ▶ vorgekommen in IrfanView 4.40

## Mögliche Reparatur

- ▶ zZ. unbekannt

# Offsetfehler

Tag 297 (PageNumber) enthält [1,1]

- ▶ erster Wert: **aktuelle Seitenzahl**, Zählung beginnt immer mit 0
- ▶ zweiter Wert: **Gesamtseitenzahl**  
(sh. TIFF-Spezifikation, S.55)
- ▶ 2 Seiten (Seite 0 und Seite 1), lt. Angabe aber nur eine
- ▶ fehlerhafte Implementierung, **kritischer Off-by-one-Fehler**
- ▶ vorgekommen in InoTec M06 Ethernet, Software ExportEngine Scanning Solution

## Mögliche Reparatur

- ▶ Löschen des Tags mit `fixit_tiff -i broken_infile.tif -x 297 -o possibly_repaired_outfile.tif`

# Offsetfehler

Tag 297 (PageNumber) enthält [1,1]

- ▶ erster Wert: **aktuelle Seitenzahl**, Zählung beginnt immer mit 0
- ▶ zweiter Wert: **Gesamtseitenzahl**  
(sh. TIFF-Spezifikation, S.55)
- ▶ 2 Seiten (Seite 0 und Seite 1), lt. Angabe aber nur eine
- ▶ fehlerhafte Implementierung, **kritischer Off-by-one-Fehler**
- ▶ vorgekommen in InoTec M06 Ethernet, Software  
ExportEngine Scanning Solution

## Mögliche Reparatur

- ▶ Löschen des Tags mit `fixit_tiff -i broken_infile.tif  
-x 297 -o possibly_repaired_outfile.tif`

# mehrere NULL-Bytes

## Tag MAKE (271) enthält mehrere NULL-Bytes

- ▶ Durch \0 darf man Strings trennen
- ▶ **aufeinanderfolgende NULL-Bytes sind aber nicht erlaubt** (sh. TIFF-Spezifikation, S. 15)
- ▶ Hinweis auf **fehlerhafte Implementierung**
- ▶ vorgekommen in Zeutschel OS1600

## Mögliche Reparatur

- ▶ mit tiffset (libtiff):

```
tiffset -s 271 "OS1600" broken_infile.tif
```

# mehrere NULL-Bytes

## Tag MAKE (271) enthält mehrere NULL-Bytes

- ▶ Durch `\0` darf man Strings trennen
- ▶ **aufeinanderfolgende NULL-Bytes sind aber nicht erlaubt** (sh. TIFF-Spezifikation, S. 15)
- ▶ Hinweis auf **fehlerhafte Implementierung**
- ▶ vorgekommen in Zeutschel OS1600

## Mögliche Reparatur

- ▶ mit `tiffset` (`libtiff`):  
`tiffset -s 271 "OS1600" broken_infile.tif`

# Zeichen außerhalb des ASCII-Charsets

Tag ImageDescription (270) enthält Umlaut 'ä'

- ▶ TIFF verwendet intern 8 Bits für Zeichensetzung
- ▶ lt. Spezifikation **dürfen aber nur Zeichen aus dem ASCII-Zeichensatz verwendet werden**  
(sh. TIFF-Spezifikation, S. 15)
- ▶ Hinweis auf **fehlerhafte Nutzereingabe**

## Mögliche Reparatur

- ▶ mit tiffset (libtiff):  
`tiffset -s 270 "Beschreibung ohne ANSI oder UTF8"  
broken_infile.tif`

# Zeichen außerhalb des ASCII-Charsets

Tag ImageDescription (270) enthält Umlaut 'ä'

- ▶ TIFF verwendet intern 8 Bits für Zeichensetzung
- ▶ lt. Spezifikation **dürfen aber nur Zeichen aus dem ASCII-Zeichensatz verwendet werden** (sh. TIFF-Spezifikation, S. 15)
- ▶ Hinweis auf **fehlerhafte Nutzereingabe**

## Mögliche Reparatur

- ▶ mit tiffset (libtiff):  
`tiffset -s 270 "Beschreibung ohne ANSI oder UTF8"  
broken_infile.tif`

# invalider DateTime-String (1)

Tag DateTime (306) enthält “Tue Dec 19 09:18:54 2006%0A”

- ▶ Software schrieb mit `localtime()` falsche Datumsangabe in Tag
- ▶ korrekt wäre: “YYYY:MM:DD HH:MM:SS\0” (sh. TIFF-Spezifikation, S. 31)
- ▶ Hinweis auf **Fehler in der Implementierung**
- ▶ vorgekommen in Accusoft ImageGear (v7.01.002)

## Mögliche Reparatur

- ▶ mit `fixit_tiff`:  

```
fixit_tiff -i broken_infile.tif -o  
possibly_repaired_outfile.tif
```

# invalider DateTime-String (1)

Tag DateTime (306) enthält "Tue Dec 19 09:18:54 2006%0A"

- ▶ Software schrieb mit `localtime()` falsche Datumsangabe in Tag
- ▶ korrekt wäre: "YYYY:MM:DD HH:MM:SS\0" (sh. TIFF-Spezifikation, S. 31)
- ▶ Hinweis auf **Fehler in der Implementierung**
- ▶ vorgekommen in Accusoft ImageGear (v7.01.002)

## Mögliche Reparatur

- ▶ mit `fixit_tiff`:  
`fixit_tiff -i broken_infile.tif -o possibly_repaired_outfile.tif`

# invalider DateTime-String (2)

Tag DateTime (306) enthält "04.03.2010 09:59:17"

- ▶ Software schrieb mit falschem Trenner Datumsangabe in Tag
- ▶ **korrekt wäre: "YYYY:MM:DD HH:MM:SS\0"**  
(sh. TIFF-Spezifikation, S. 31)
- ▶ Hinweis auf **Fehler in der Implementierung**
- ▶ vorgekommen in PROView (Business Graphics Datentechnik GmbH, 2010)

## Mögliche Reparatur

- ▶ mit `fixit_tiff`:  
`fixit_tiff -i broken_infile.tif -o  
possibly_repaired_outfile.tif`

# invalider DateTime-String (2)

Tag DateTime (306) enthält “04.03.2010 09:59:17”

- ▶ Software schrieb mit falschem Trenner Datumsangabe in Tag
- ▶ **korrekt wäre: “YYYY:MM:DD HH:MM:SS\0”**  
(sh. TIFF-Spezifikation, S. 31)
- ▶ Hinweis auf **Fehler in der Implementierung**
- ▶ vorgekommen in PROView (Business Graphics Datentechnik GmbH, 2010)

## Mögliche Reparatur

- ▶ mit `fixit_tiff`:  
`fixit_tiff -i broken_infile.tif -o possibly_repaired_outfile.tif`

# Private Tags außerhalb IFD

## EXIF-Daten

- ▶ **Private Tags, die en block bestellt sind, sollten privates IFD nutzen** (sh. TIFF-Spezifikation, S. 9)
- ▶ EXIF-Tag 36867 (DateTimeOriginal) wird direkt im ersten IFD des TIFF referenziert, nicht über EXIF-IFD
- ▶ Hinweis auf **ungenügende Spezifikation**
- ▶ vorgekommen in ProServ ScannTech ST 602i-3, Software PROView 7.99

## Mögliche Reparatur

- ▶ Löschen des Tags mit `fixit_tiff -i broken_infile.tif -x 36867 -o possibly_repaired_outfile.tif`

# Private Tags außerhalb IFD

## EXIF-Daten

- ▶ **Private Tags, die en block bestellt sind, sollten privates IFD nutzen** (sh. TIFF-Spezifikation, S. 9)
- ▶ EXIF-Tag 36867 (DateTimeOriginal) wird direkt im ersten IFD des TIFF referenziert, nicht über EXIF-IFD
- ▶ Hinweis auf **ungenügende Spezifikation**
- ▶ vorgekommen in ProServ ScannTech ST 602i-3, Software PROView 7.99

## Mögliche Reparatur

- ▶ Löschen des Tags mit `fixit_tiff -i broken_infile.tif -x 36867 -o possibly_repaired_outfile.tif`

# Überblick

Vor- und Nachteile TIFF

Problemfälle

Werkzeuge

Zusatz: checkit\_tiff

# Werkzeuge Validierung - Jhove

```
$> jhove -m TIFF-hul minimal_valid.tiff
```

```
Jhove (Rel. 1.6, 2011-01-04)
Date: 2015-09-14 16:19:36 MESZ
RepresentationInformation: minimal_valid.tiff
ReportingModule: TIFF-hul, Rel. 1.5 (2007-10-02)
LastModified: 2015-09-10 15:04:48 MESZ
Size: 323
Format: TIFF
Version: 5.0
Status: Well-Formed and valid
SignatureMatches:
  TIFF-hul
MIMEtype: image/tiff
Profile: Baseline bilevel (Class B), TIFF/IT-BP (ISO 12639:1996), TIFF/IT-MP (ISO 12639:1996)
TIFFMetadata:
  ByteOrder: little-endian
  IFDs:
    Number: 1
    IFD:
      Offset: 38
      Type: TIFF
      Entries:
        NisoImageMetadata:
          ByteOrder: little_endian
          CompressionScheme: uncompressed
          ImageWidth: 20
          ImageHeight: 10
          ColorSpace: white is zero
          Orientation: normal
          SamplingFrequencyUnit: inch
          XSamplingFrequency: 376,193
          YSamplingFrequency: 376,193
          BitsPerSample: 1
          BitsPerSampleUnit: integer
          SamplesPerPixel: 1
        NewSubfileType: 0
        DocumentName: /home/romeyke/fixit_tiff/checkit/tiffs_should_pass/minimal_valid2.tiff
        SampleFormat: 1
        MinSampleValue: 0
        MaxSampleValue: 1
        Thresholding: 1
      TIFFITProperties:
        BackgroundColorIndicator: background not defined
        ImageColorIndicator: image not defined
        TransparencyIndicator: no transparency
        PixelIntensityRange: 0, 1
        RasterPadding: 1 byte
        BitsPerRunLength: 8
        BitsPerExtendedRunLength: 16
```

# Werkzeuge Validierung - tiffinfo (libtiff)

```
$> tiffinfo minimal_valid.tiff
```

```
TIFF Directory at offset 0x26 (38)
```

```
Subfile Type: (0 = 0x0)
```

```
Image Width: 20 Image Length: 10
```

```
Resolution: 376.193, 376.193 pixels/inch
```

```
Bits/Sample: 1
```

```
Compression Scheme: None
```

```
Photometric Interpretation: min-is-white
```

```
Orientation: row 0 top, col 0 lhs
```

```
Samples/Pixel: 1
```

```
Rows/Strip: 64
```

```
Planar Configuration: single image plane
```

```
DocumentName: /home/romeyke/fixit\_tiff/checkit/tiffs_should_pass/minimal_valid2.tiff
```

# Werkzeuge Validierung - checkit\_tiff

```
$> checkit_tiff minimal_valid.tiff  
cit_tiff6_baseline_SLUB.cfg
```

```
'checkit_tiff' version: 0.228_g9766064-dirty,  
  licensed under conditions of libtiff (see http://libtiff.maptools.org/misc.html)  
tiff file=minimal_valid.tiff  
cfg_file=cit_tiff6_baseline_SLUB.cfg  
check if all IFDs are word aligned  
check if only one IFD exists  
check if tags are in ascending order  
check if all offsets are used once only  
check if all offsets are word aligned  
check if tag 256 (ImageWidth) has value in range 1 - 4294967295  
check if tag 257 (ImageLength) has value in range 1 - 4294967295  
check if tag 258 (BitsPerSample) has some of these 3-values, 8, 4, 1  
check if tag 259 (Compression) has value 1  
check if tag 262 (PhotometricInterpretation) has value in range 0 - 2  
check if tag 273 (StripOffsets) exists  
check if tag 273 (StripOffsets) has valid type  
check if tag 273 (StripOffsets) has valid asciivalue  
check if tag 277 (SamplesPerPixel) has value 1  
check if tag 278 (RowsPerStrip) has value in range 1 - 4294967295  
check if tag 279 (StripByteCounts) has value in range 1 - 4294967295  
check if tag 282 (XResolution) has value in range 300 - 1200  
check if tag 283 (YResolution) has value in range 300 - 1200  
check if tag 296 (ResolutionUnit) has value 2  
check if tag 254 (SubfileType) has value 0  
check if tag 269 (DocumentName) has value matching regex '^[[[:print:]]]*$'  
check if tag 274 (Orientation) has value 1  
check if tag 284 (PlanarConfiguration) has value 1  
check if forbidden tags are still existing  
the given tiff is valid
```

# Werkzeuge Validierung - DPFManager 1.4

## Validierungsansicht, einzelne Datei

CONFIRMANCE CHECKER | REPORTS | ABOUT

**DPF Manager**  
MULTIFILE REPORT

**1 file processed**

1 conforms to policy checker  
2 conforms to license checker

0 passed  
1 failed  
Global score 0%

invalid\_data.txt  
Conformance checker  
Baseline 1 errors 0 warnings

Failed  
Score 87%

# Quellen Validierungswerkzeuge

- ▶ JHove (<http://openpreservation.org/technology/products/jhove/>)
- ▶ libtiff (<http://libtiff.maptools.org>)
- ▶ **checkit\_tiff** ([https://github.com/SLUB-digitalpreservation/checkit\\_tiff](https://github.com/SLUB-digitalpreservation/checkit_tiff))
- ▶ DPFManager der TIFF/A - Initiative (<http://www.dpfmanager.org/>)

# Werkzeuge zur Reparatur

- ▶ libtiff
  - ▶ <http://libtiff.maptools.org>
  - ▶ CLI, jede Plattform, OpenSource, verwendet in SLUB
- ▶ fixit\_tiff
  - ▶ [https://github.com/SLUB-digitalpreservation/fixit\\_tiff](https://github.com/SLUB-digitalpreservation/fixit_tiff)
  - ▶ CLI, Linux/AIX/Windows, OpenSource, verwendet in SLUB
- ▶ ~~Tiff Tag Editor~~
  - ▶ <http://www.leadtools.com>
  - ▶ GUI, Windows, proprietär, von uns nicht getestet, eingestellt
- ▶ Advanced TIFF Editor
  - ▶ <http://www.tiffedit.com>
  - ▶ GUI, Windows, proprietär, von uns nicht getestet

# Was immer geht

- ▶ Zettel und Bleistift,
- ▶ Taschenrechner
- ▶ Hex-Editor eigener Wahl

# Danke an

- ▶ meine Kollegen Jörg Sachse und Jens Steidl für Diskussionen zur Interpretation der TIFF-Spezifikation und Sammlung der Problemfälle
- ▶ meinen Kollegen Henning Gerhardt, der das `checkit_tiff` -Tool unter die Lupe nahm
- ▶ Ange Albertini für <http://pics.corkami.com> und etliche Anregungen
- ▶ libtiff-Entwickler und jHove-Entwickler für wertvolle Hinweise

# Überblick

Vor- und Nachteile TIFF

Problemfälle

Werkzeuge

Zusatz: checkit\_tiff

# checkit\_tiff: Beispielkonfiguration baseline TIFF

```
# tag; required; values
#
# works as whitelist
# * tag: any tag number in TIFF
# * required: mandatory | optional | depends( $tag,$value )
# * values: range($start, $end) | logical_or( $a, _) | only($a) | any

# This defines a sample config for baseline tiffs
# Remember, because it works as whitelist for each required tag we need an
# entry here

# Required Baseline Tags

# 256 0100 ImageWidth The number of columns in the image, i.e., the number of pixels per row.
256; mandatory; any

# 257 0101 ImageLength The number of rows of pixels in the image.
257; mandatory; any

# 258 0102 BitsPerSample Number of bits per component.
258; mandatory; logical_or(8,16)

# 259 0103 Compression Compression scheme used on the image data
### (1 means no compression)
259; mandatory; only(1)

# 262 0106 PhotometricInterpretation The color space of the image data.

### 2 means RGB, 0 and 1 means bilevel or grayscale, 0 is unusual
262; mandatory; logical_or(2,1)

# 273 0111 StripOffsets For each strip, the byte offset of that strip.
273; mandatory; any

# 277 0115 SamplesPerPixel The number of components per pixel.
### if RGB then 3 else 1
277; depends(262.2); only(3)
277; depends(262.1); only(1)

# 278 0116 RowsPerStrip The number of rows per strip.
278; mandatory; any

# 279 0117 StripByteCounts For each strip, the number of bytes in the strip after compression.
279; mandatory; any

# 282 011A XResolution The number of pixels per ResolutionUnit in the ImageWidth direction.
282; mandatory; range(300, 1200)

# 283 011B YResolution The number of pixels per ResolutionUnit in the ImageLength direction.
283; mandatory; range(300, 1200)

# 296 0128 ResolutionUnit The unit of measurement for XResolution and YResolution.
### (2 means inch)
296; mandatory; only(2)
```