

TIFF Konvertierung *Out of Sequence und Type Mismatch*

1	Vorwort	1
2	Ausgangslage	1
3	Analyse	3
	3.1 Out of sequence	3
	3.2 Type mismatch	4
4	Konvertierung	5
	4.1 Konvertierungsbedarf	5
	4.2 Konvertierungsanleitung	5

1 Vorwort

Diese Konvertierungsanleitung soll betroffenen Archiven beispielhaft aufzeigen, wie sie allgemein mit TIFF-Dateien, die die Validierung mit KOST-Val¹ nicht bestanden haben, und im speziellen mit den Fehlern "out of sequence" und "Type mismatch" umgehen sollen.

Bei invaliden TIFF-Dateien wird empfohlen, als erstes Art und Häufigkeit des Fehlers zu analysieren. Handelt es sich wirklich um einen Fehler in der TIFF-Datei, ist der Konvertierungsbedarf zu ermitteln. Falls eine Konvertierung angezeigt ist, muss diese zuvor getestet und überprüft und im Anschluss dokumentiert werden. Wenn mehrere TIFFs mit dem gleichen Fehler aus der gleichen Quelle stammen, sollte zusätzlich zu einer eventuellen Konvertierung überprüft werden, ob die Einstellungen dieser Quelle so angepasst werden können, dass der Fehler nicht weiter erzeugt wird.

2 Ausgangslage

Einige TIFF-Dateien haben bei der TIFF-Validierung durch KOST-Val die Module B bis G nicht bestanden. Der kost-val.log zeigt jedoch auf, dass insbesondere das Modul B (Jhove mit der Fehlermeldung: "out of sequence" und "Type mismatch") invalid ist und die restlichen Module aufgrund der misslungenen Jhove-Validierung nicht bestanden haben.

¹ Der KOST-Validator für TIFF- und andere Dateien – KOST-Val – führt nach einer vorgängigen TIFF-Erkennung die Jhove-Validierung durch. Anschliessend wird der Jhove-Report für die detailliertere Analyse der Tags im Bereich Komprimierung, Farbraum, BitsPerSample, Multipage, Kacheln und Grösse verwendet. Mit dem KOST-Val kann validiert werden, ob ein TIFF der TIFF-Präzisierung entspricht oder nicht.

```
TIFF Validierung -> Digitalisat.tif

[A] = Valid
[B] = Invalid (Test "Jhove" nicht bestanden)
[B] -- Die Jhove-Validierung wurde nicht bestanden ( Status: Not well-formed).
[B] -- ErrorMessage: Tag 34377 out of sequence
[B] -- ErrorMessage: Tag 33723 out of sequence
[B] -- ErrorMessage: Tag 270 out of sequence
[B] -- ErrorMessage: Type mismatch for tag 270; expecting 2, saw 1
[C] = Invalid (Test "Komprimierung" nicht bestanden)
[C] -- Die Validierung ist nicht bestanden da Jhove (Module B) das Tag nicht lesen konnte.
[D] = Invalid (Test "Farbraum" nicht bestanden)
[D] -- Die Validierung ist nicht bestanden da Jhove (Module B) das Tag nicht lesen konnte.
[E] = Invalid (Test "BitsPerSample" nicht bestanden)
[E] -- Die Validierung ist nicht bestanden da Jhove (Module B) das Tag nicht lesen konnte.
[F] = Invalid (Test "Multipage" nicht bestanden)
[F] -- Die Validierung ist nicht bestanden da Jhove (Module B) das Tag nicht lesen konnte.
[G] = Invalid (Test "Kacheln" nicht bestanden)
[G] -- Die Validierung ist nicht bestanden da Jhove (Module B) das Tag nicht lesen konnte.
[H] = Valid

TOTAL = Invalid
```

Die Fehler aus dem jhove.log.txt sind direkt im kost-val.log ersichtlich. Öffnet man zusätzlich auch den jhove-log.txt², sieht man, dass keine Tag-Information gelesen wurde respektive gelesen werden konnte.

```
Jhove (Rel. 1.5, 2009-12-19)
Date: 2014-03-03 11:26:35 MEZ
RepresentationInformation: O:\AMT\Bild_OutOfSequence_TypeMismatch\Digitalisat.tif
ReportingModule: TIFF-hul, Rel. 1.5 (2007-10-02)
LastModified: 2007-06-29 15:44:52 MESZ
Size: 107594740
Format: TIFF
Status: Not well-formed
SignatureMatches:
  TIFF-hul
ErrorMessage: Tag 34377 out of sequence
  Offset: 107594688
ErrorMessage: Tag 33723 out of sequence
  Offset: 107594700
ErrorMessage: Tag 270 out of sequence
  Offset: 107594712
ErrorMessage: Type mismatch for tag 270; expecting 2, saw 1
MIMEtype: image/tiff
```

² Der Jhove-Report ist nur ersichtlich, wenn die Option Verbose [-v] gewählt wurde.

3 Analyse

Die Analyse der kost-val.log-Datei "Digitalisat.tif.kost-val.log" zeigt auf, dass Modul B "Jhove" aus 2 Hauptgründen nicht bestanden wurde. Diese zwei Fehler führten zu einem Abbruch der Jhove-Validierung; entsprechend konnten die Module C-G nicht validiert werden.

Nachfolgend werden die zwei Hauptgründe für die Invalidität der TIFF-Datei analysiert.

3.1 Out of sequence

Auszug aus dem kost-val.log mit der entsprechenden Jhove Fehlermeldung:

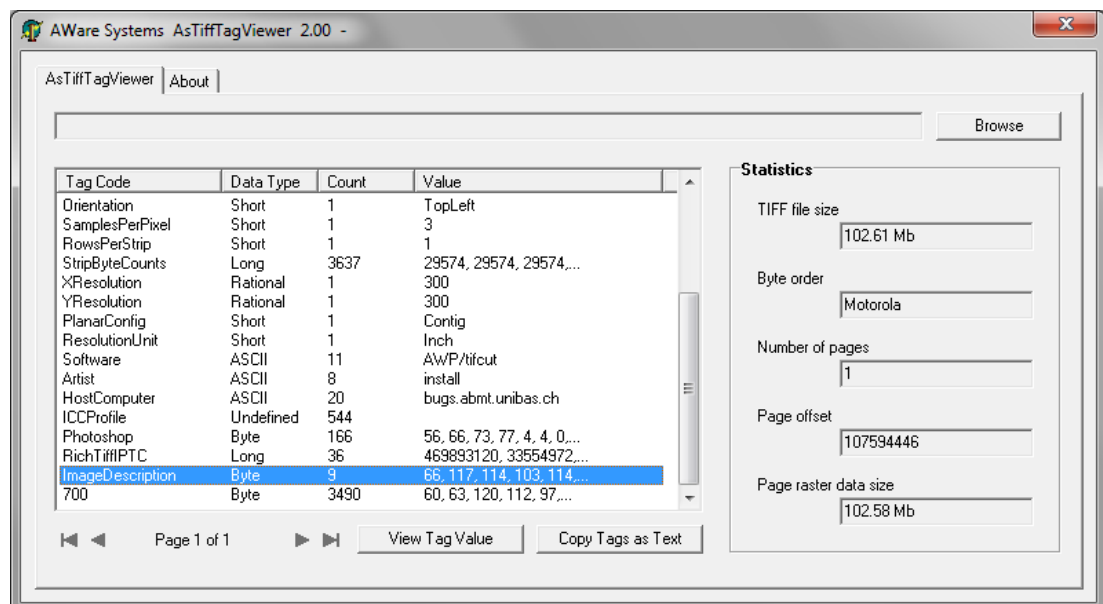
```
[B] -- ErrorMessage: Tag 34377 out of sequence
[B] -- ErrorMessage: Tag 33723 out of sequence
[B] -- ErrorMessage: Tag 270 out of sequence
```

Jhove bemängelt in diesem Fall, dass Tag 34377, 33723 und 270 "out of sequence" (nicht in der richtigen Reihenfolge) sind. Dabei validiert Jhove die Anforderung "Sort Order" aus der TIFF Spezifikation Version 6 auf der Seite 15, welche folgendes definiert:

Sort Order

The entries in an IFD must be sorted in ascending order by Tag. Note that this is not the order in which the fields are described in this document. The Values to which directory entries point need not be in any particular order in the file.

Eine Sichtung der Tags mit dem Programm AsTiffTagViewer bestätigt, dass ImageDescription (270), IPTC/NAA (33723) und PhotoshopImageResources (34377) nicht in aufsteigender Reihenfolge enthalten sind (z.B. müsste ImageDescription (270) vor Orientation (274) aufgelistet sein):



Dieser Fehler ist dem Bereich der Metadaten zuzuordnen. Die eigentlichen Bilddaten sind nicht betroffen. Entsprechend muss darauf geachtet werden, dass sich das eigentliche Bild nicht verändert, wenn die Metadaten korrigiert werden.

3.2 Type mismatch

Auszug aus dem Jhove Report mit der entsprechenden Fehlermeldung:

```
[B] -- ErrorMessage: Type mismatch for tag 270; expecting 2, saw 1
```

Jhove bemängelt in diesem Fall, dass Tag 270 nicht den richtigen Type enthält (1 statt 2). Dabei validiert Jhove die Anforderung "ImageDescription" aus der TIFF Spezifikation Version 6 auf der Seite 34. ImageDescription entspricht dem Tag 270 und ist wie folgt definiert:

ImageDescription

A string that describes the subject of the image.

Tag = 270 (10E.H)

Type = ASCII

For example, a user may wish to attach a comment such as "1988 company picnic" to an image.

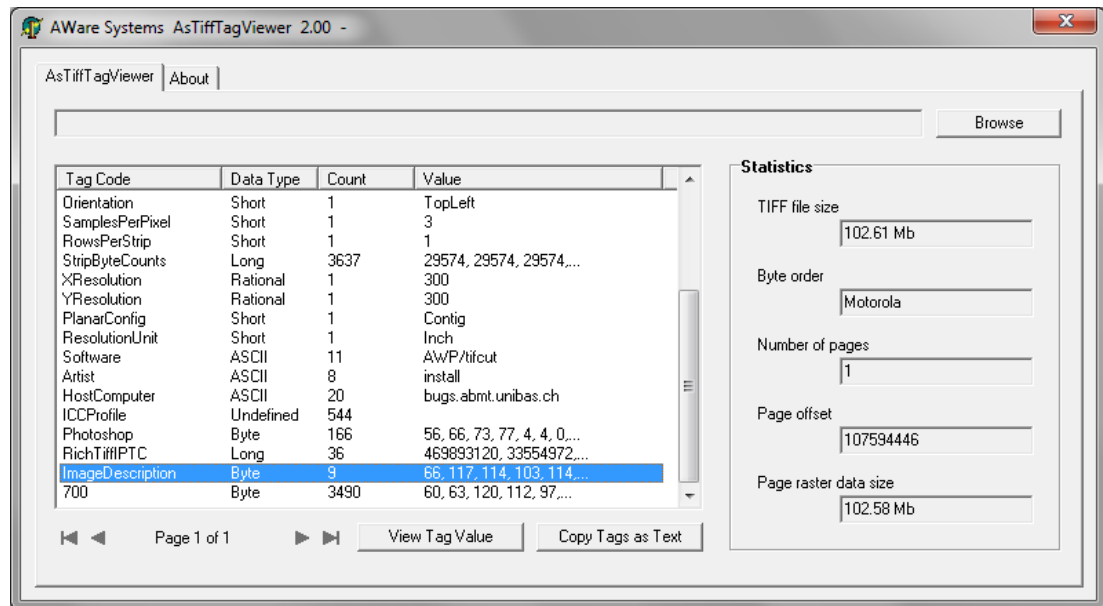
Die Spezifikation schreibt vor, dass der Type "ASCII" sein muss. Auf den Seiten 15 und 16 der Spezifikation sind die Typen folgendermassen definiert:

Types

The field types and their sizes are:

1 = BYTE	8-bit unsigned integer.
2 = ASCII	8-bit byte that contains a 7-bit ASCII code; the last byte must be NUL (binary zero).
3 = SHORT	16-bit (2-byte) unsigned integer.
4 = LONG	32-bit (4-byte) unsigned integer.
5 = RATIONAL	Two LONGs: the first represents the numerator of a fraction; the second, the denominator.
6 = SBYTE	An 8-bit signed (twos-complement) integer.
7 = UNDEFINED	An 8-bit byte that may contain anything, depending on the definition of the field.
8 = SSHORT	A 16-bit (2-byte) signed (twos-complement) integer.
9 = SLONG	A 32-bit (4-byte) signed (twos-complement) integer.
10 = SRATIONAL	Two SLONG's: the first represents the numerator of a fraction, the second the denominator.
11 = FLOAT	Single precision (4-byte) IEEE format.
12 = DOUBLE	Double precision (8-byte) IEEE format.

Jhove bemängelt entsprechend, dass der beschreibende Text im Tag 270 der TIFF-Datei vom Typ BYTE statt ASCII ist. Eine Sichtung der Tags mit dem Programm AsTiffTagViewer bestätigt dies:



Dieser Fehler ist dem Bereich der Metadaten zuzuordnen. Die eigentlichen Bilddaten sind nicht betroffen. Entsprechend muss darauf geachtet werden, dass sich das eigentliche Bild nicht verändert, wenn die Metadaten korrigiert werden.

4 Konvertierung

4.1 Konvertierungsbedarf

Solange ein TIFF nicht valid ist, erhöht sich das Risiko, dass diese Datei in Zukunft nicht konvertiert werden kann oder das Metadaten verloren gehen.

Die Konvertierungsempfehlung der KOST Preservation Planning Expertengruppe empfiehlt bei mehreren nicht empfohlenen Eigenschaften, im vorliegenden Fall zwei Fehlertypen in den Metadaten, die TIFFs möglichst rasch zu konvertieren.

4.2 Konvertierungsanleitung

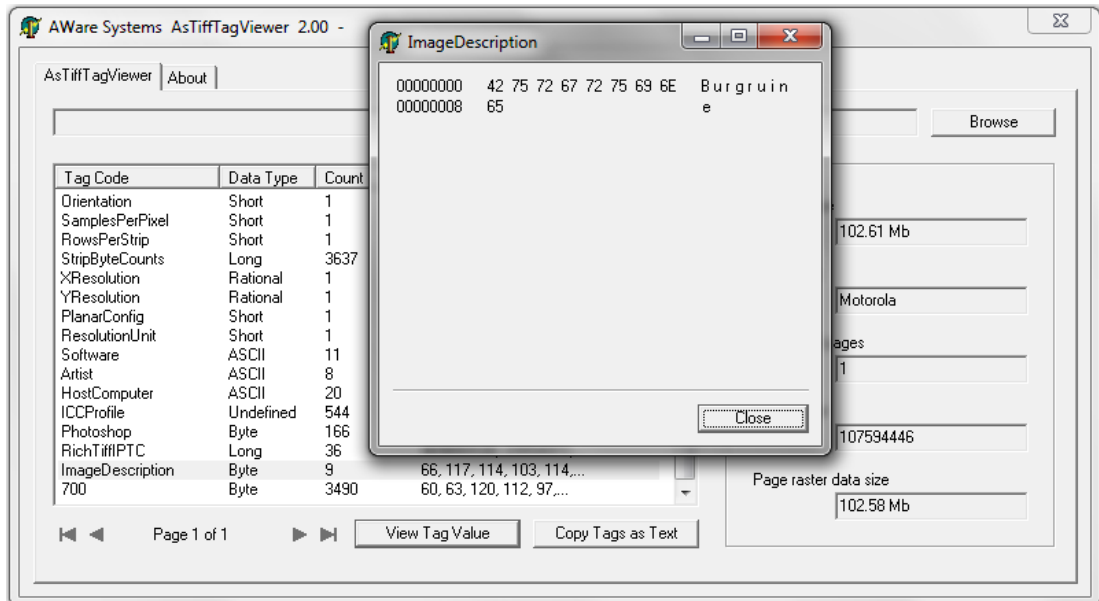
Bei diesen Fehlern sind nur Metadaten betroffen. Entsprechend müssen bei der Konvertierung die Bilddaten sowie die von den Fehlern nicht betroffenen Metadaten unverändert bleiben. Dabei muss beachtet werden, dass sich aber die Offset-Werte zu den einzelnen Streifen verändern können, da die TIFF-Datei neu zusammengestellt wird. Diese Anforderungen erfüllen gemäss unseren Kenntnissen nur "ExifToolGUI" von Bogdan Hrastnik mit dem zugrundeliegenden "ExifTool" von Phil Harvey sowie das reine "ExifTool" als Commandline-Programm.

Nachfolgend wird die TIFF-Datei mit dem "ExifToolGUI" korrigiert. Das GUI ist verständlicher und erlaubt durch Selektion von verschiedenen Dateien auch eine Stapelverarbeitung. Die Stapelverarbeitung kann jedoch nur verwendet werden, wenn der neue Wert immer der gleiche ist.

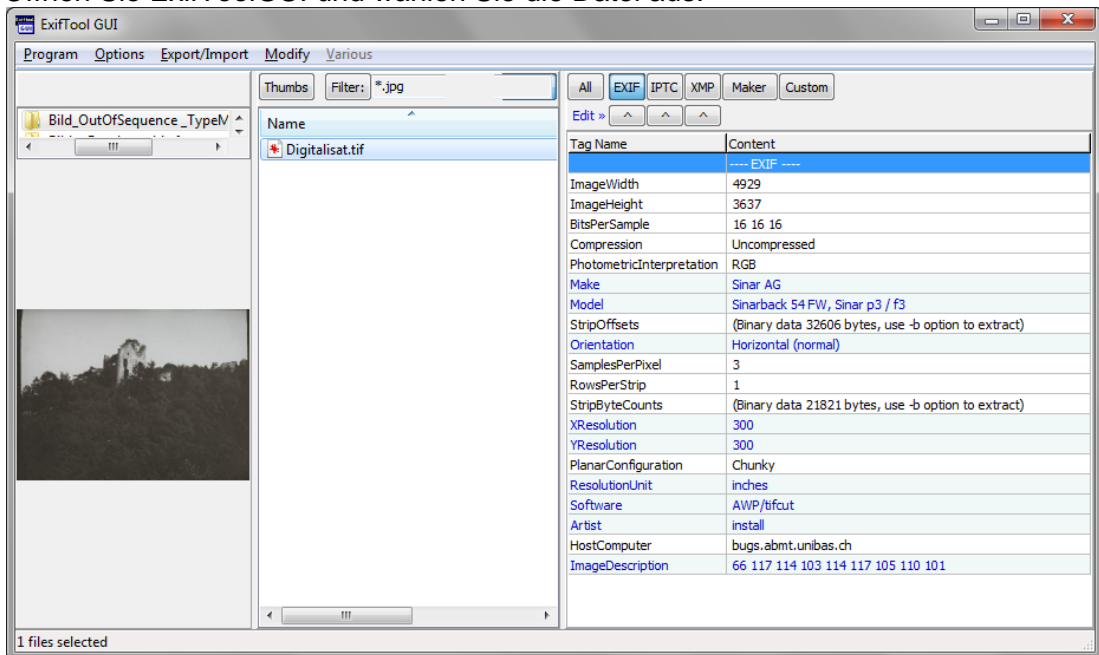
4.2.1 Konvertierung mit ExifToolGUI inkl. Kontrolle mit TIFF-Val

1. Laden Sie ExifToolGUI und ExifTool herunter und installieren Sie beide Programme nach der Anleitung (<http://u88.n24.queensu.ca/~bogdan/>).

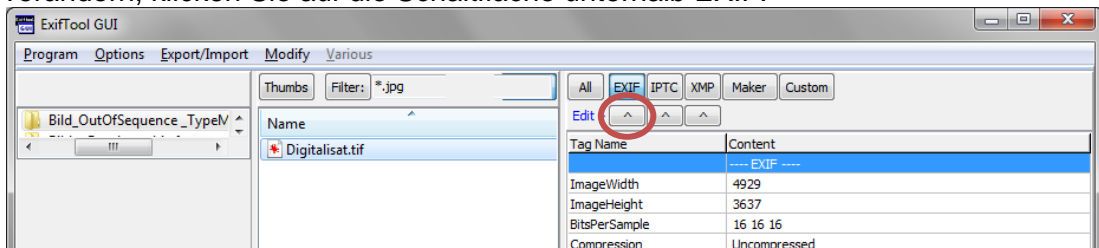
- Als nächstes müssen Sie wissen, welche Metadaten mit welchen Werten zu ändern sind. In unserem Beispiel ist dies ImageDescription, welches den Text "Burgruine" in ASCII enthalten soll.



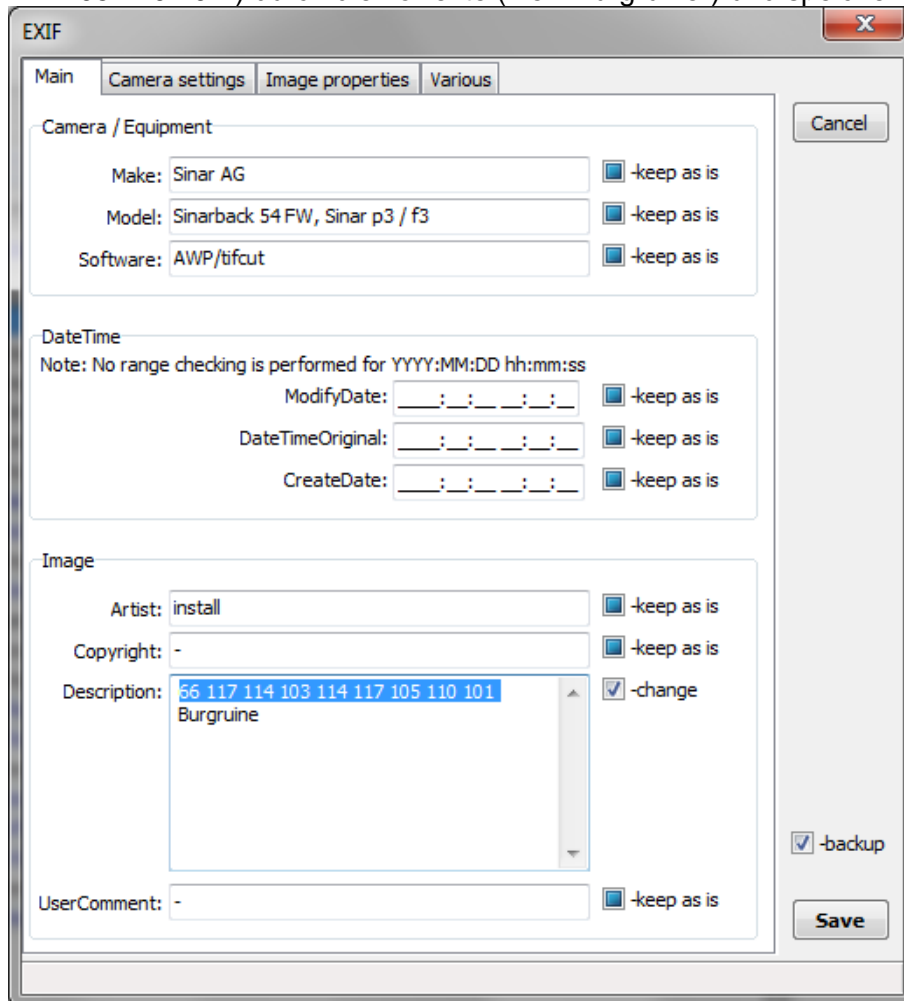
- Öffnen Sie ExifToolGUI und wählen Sie die Datei aus.



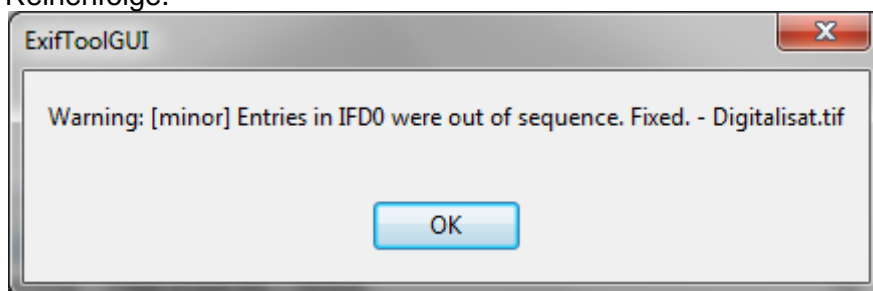
- ImageDescription ist ein normales TIFF-Tag und erscheint im ExifToolGUI unter EXIF (zuerst in der Liste, da die Reihenfolge nicht stimmt). Um die EXIF-Werte zu verändern, klicken Sie auf die Schaltfläche unterhalb EXIF.



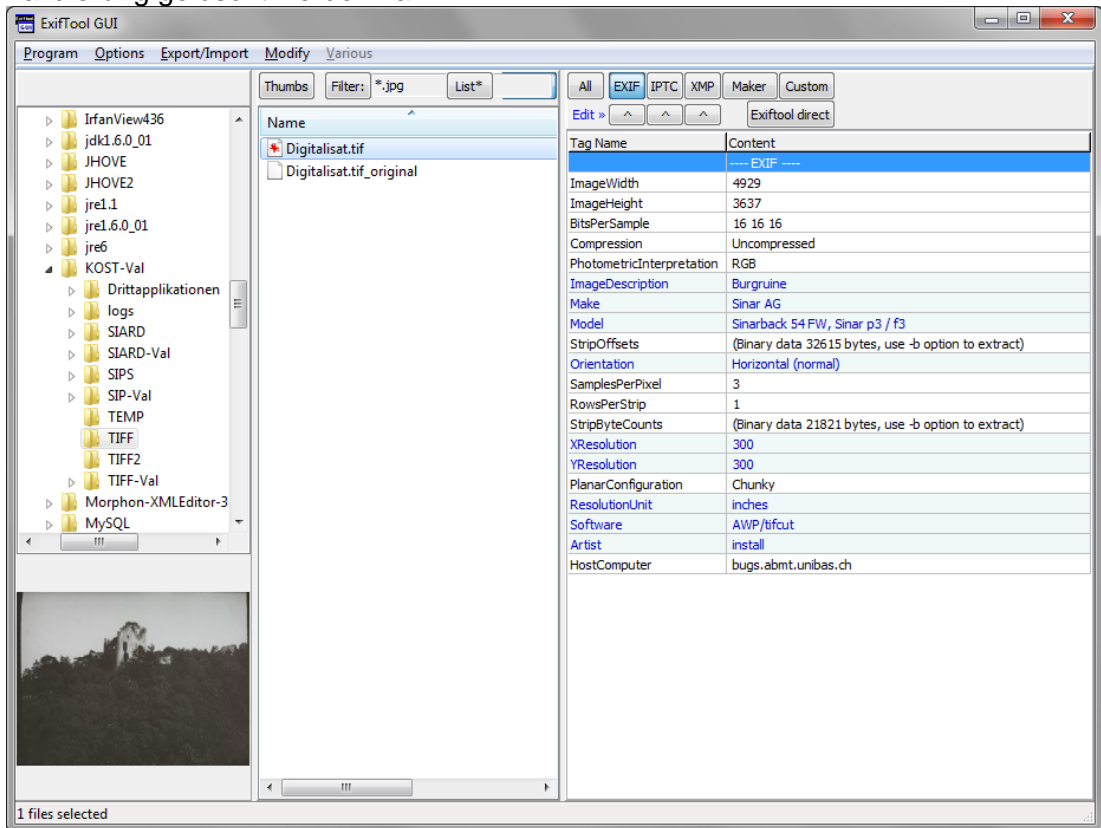
5. Ersetzen Sie im neuen Fenster die fehlerhafte Description (hier "66 117 114 103 114 117 105 110 101") durch die korrekte (hier "Burgruine") und speichern Sie.



6. Beim Speichern stellt ExifTool die TIFF-Datei (genauer die einzelnen Metadaten/Tags) neu zusammen und bereinigt gleichzeitig die fehlerhafte Reihenfolge.



7. Jetzt ist ImageDescription lesbar, und auch die Reihenfolge stimmt. ExifTool erstellt zudem eine Kopie der Originaldatei(*.tif_original), welche nach erfolgreicher Validierung gelöscht werden kann.



8. Validieren Sie die Datei erneut mit KOST-Val. Es ist grundsätzlich möglich, dass die konvertierte Datei ein weiteres Mal konvertiert werden muss, da bis jetzt die Validierungsschritte C bis G aufgrund des Jhove-Absturzes nicht durchgeführt werden konnte. Die Validierung ist in unserem Beispiel jetzt bestanden.

