DPF MANAGER

TIFF/A STANDARD PROPOSAL

Project acronym: PREFORMA
PREFORMA - Future Memory Standards
PREservation FORMAts for culture information/e-archives
EC Grant agreement no: 619568
EC Call ID: FP7-ICT-2013-11

TIFF/A

Tagged Image File Format for Archival - Draft Specification

February 2015

University of Basel, Switzerland

UNI BASEL

University of Girona, Spain

Universitat de Girona

Easy Innova

easy

Copyright notice

University of Basel
Digital Humanities Lab
Bernoullistrasse 32
4056 Basel
Switzerland
peter.fornaro@unibas.ch
+41612670487

University of Basel, Switzerland

University of Girona, Spain

Easy Innova

1. Contents

2. Introduction

Any file format basically defines the *logical structure and meaning* of the bits within the bitstream and thus it is essential for correct interpretation and proper rendering of the coded data. Unfortunately a file format or parts of its logical structure and definition can become obsolete. As a result the information renders useless, even-though the bit stream is still properly readable. Therefore it is necessary to use a stable and proper file format for long-term preservation of digital data. The stability of a format is determined by the following criteria:

- the format has to be well documented
- the format should be in wide used
- the format should not contain proprietary or patented elements (algorithms etc.). In other words, it should be an *open* format.
- the format should be as simple as possible
- no interlinkage with external data is allowed (e.g. fonts or any other resource)
- In addition to that it must be given, that the file-format is capable to store the relevant information without significant compromises. In the case of image files this could for

example be the quantisation depth for correct tonal reproduction, that shall be bigger than 8bit/channel for most applications.

TIFF is the most widely used and most accepted format for this task. However, TIFF itself is quite complex and the different standards TIFF V6.0, TIFF/EP (ISO 12234-2:2001) etc. include many options that are not suitable for long term archival. Therefore it might not be possible to render a file in future due to the lack of documentation. To overcome this problem a limitation of functionalities of TIFF- files can be a suitable solution. Therefore a set of mandatory, optional and unsuitable options are defined in order to guarantee that in future an image file can be rendered without loss of quality and information, even if hardware and software will have changed significantly.
We therefore propose a subset of TIFF which is fully compatible with the TIFF standard but marks some tags as mandatory, some as optional and some as forbidden in order to guarantee the correct rendering in the future. In analogy to PDF/A format we propose to call this Format TIFF/Archive or TIFF/A.

3. Scope

This part of ISO xxx defines the TIFF/A data format, the Tagged Image File Format for archival use. The Format is a derivative of TIFF by the means that the format is shaped for sustainability that only is given if all information is valid and well documented. Therefor all non standard information is forbidden.

3.1. Normative Reference

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

IEC 61966-2-1:1999, Multimedia Systems and Equipment — Colour Measurement and Management, Part 2-1: Colour management — Default RGB Colour Space — sRGB

ICC.1:2001-12, File Format for Color Profiles, International Color Consortium

TIFF, Revision 6.0 Final, Aldus Corporation (now Adobe Systems Incorporated), June 3, 1992

4. Terms, definitions and abbreviated terms

For the purposes of this International Standard, the following definitions apply:

4.1. File system

Software structure which specifies how the data is logically organized on a given storage media.

4.2. Image data format

Structure and content which specifies how the data is logically organized on a given storage media.

5. Image data features

This clause describes all the features of the TIFF/A standard and lists the tags used to implement each feature.

5.1. TIFF/A file encoding structure

A TIFF/A file is a valid TIFF file that contains the TIFF/A format identifier and conforms to the restrictions described in this document. The TIFF/A header is exactly the same as the TIFF header. The use of the TIFF/A format and revision number is identified in the TIFF/A StandardID tag-field.
TIFF is an image file format. In this document, a file is a sequence of 8-bits bytes, where the bytes are numbered from 0 to N. The largest possible TIFF file is 2**32 bytes in length. A TIFF file begins with an 8-byte image file header that points to an image file directory (IFD). An image file directory contains information about the image, as well as pointers to the actual image data.

A TIFF file structure is shown in Figure 1.

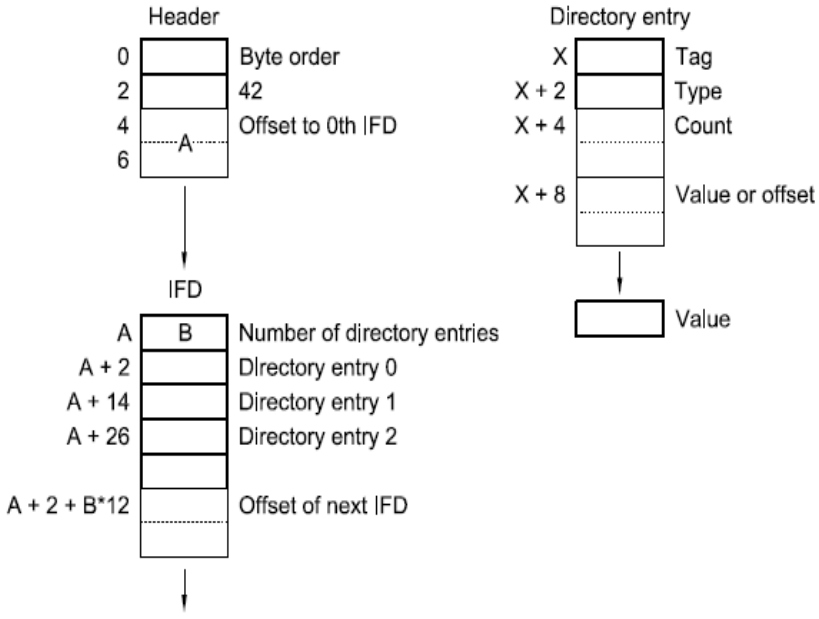

Figure 1 — TIFF file structure

5.1.1. Image file header

A TIFF file begins with an 8-byte image file header, containing the following information:

Bytes 0-1: The byte order used within the file. Legal values are:
II (4949.H)
MM (4D4D.H)
In the II format, byte order is always from the least significant byte to the most significant byte, for both 16-bit and 32-bit integers. This is called little-endian byte order. In the MM format, byte order is always from the most significant byte to the least significant byte, for both 16-bit and 32-bit integers. This is called big-endian byte order.

Bytes 2-3: An arbitrary but carefully chosen number (42) that further identifies the file as a TIFF file.
The byte order depends on the value of Bytes 0-1. Note that, in order to allow backward compatibility with future versions of TIFF/A based on future
versions of TIFF, all TIFF/A readers should test the TIFF header version value to determine if VERSION >= 42, not to test if VERSION == 42. This will allow higher version numbers to be used in the future and to be detected as TIFF/A files.

Bytes 4-7: The offset (in bytes) of the first IFD. The directory may be at any location in the file after the header but shall begin on a word boundary. In particular, an image file directory may follow the image data it describes. Readers shall follow the pointers wherever they may lead.
The term byte offset is always used in this part of ISO 12234 to refer to a location with respect to the beginning of the TIFF file. The first byte of the file has an offset of 0.

5.1.2. Image file directory

An IFD consists of a 2-byte count of the number of directory entries (i.e. the number of fields), followed by a sequence of 12-byte field entries, followed by a 4-byte offset of the next IFD (or 0 if none). Do not forget to write 4 bytes of 0 after the last IFD.
There shall be at least 1 IFD in a TIFF file and each IFD shall have at least one entry.

IFD entry
Each 12-byte IFD entry has the following format:
Bytes 0-1: The tag that identifies the field.
Bytes 2-3: The field type.
Bytes 4-7: The number of values, Count of the indicated type.
Bytes 8-11: The value offset, i.e. the file offset (in bytes) to the value(s) for the field. This value offset is expected to begin on a word boundary; the corresponding value offset will thus be an even number. This file offset may point anywhere in the file, even after the image data. (See below for more information.)

IFD terminology

A TIFF field is a logical entity consisting of a TIFF tag and its value. This logical concept is implemented as an IFD entry, plus the actual value if it doesn't fit into the value/offset part, the last 4 bytes of the IFD entry. The terms TIFF field and IFD entry are interchangeable in most contexts.

IFD sort order
The entries in an IFD shall be sorted in ascending order by tag. The values to which directory entries point need not be in any particular order in the file.

Value/Offset
To save time and space, the value offset contains the value instead of pointing to the value only if the value fits into 4 bytes. If the value is shorter than 4 bytes, it is left-justified within the 4-byte value offset, i.e. stored in the lower numbered bytes. Whether the value fits within 4 bytes is determined by the type and count of the field.
Note that the 4 byte value offset should not be thought of as a LONG data type since, if the value is shorter than 4 bytes, it is always left-justified regardless of whether the II or MM byte order is used. For example, to store the SHORT hex value "AB CD" in MM byte order, the 4 bytes are "AB CD xx xx" (where x indicates "don't care"). The same hex value in II byte order is given by "CD AB xx xx".

Count
Count, called Length in previous versions of the TIFF specification, is the number of values. Note that count is not the total number of bytes. For example, a single 16-bit word (SHORT) has a count of 1 not 2.

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 shall 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 represents 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 SLONGs: the first represents the numerator of the fraction, the second represents the denominator.
11=FLOAT Single precision (4-byte) IEEE format.
12=DOUBLE Double precision (8-byte) IEEE format.

WARNING — It is possible that other TIFF field types will be added in the future. Readers should ignore fields containing an unexpected field type.

The value of the count part of an ASCII field entry includes the NULL. If padding is necessary, the count does not include the pad byte. Note that there is no initial count byte as in Pascal-style strings. Any ASCII field can contain multiple strings, each terminated with a NUL. A single string is preferred whenever possible. The count for multistring fields is the number of bytes in all strings in that field plus their terminating NUL bytes. Only one NUL is allowed between strings, so that the strings following the first string will often begin on an odd byte.

The reader shall check the type to verify that it contains an expected value. TIFF currently allows more than 1 valid type for some fields. For example, ImageWidth and ImageLength are usually specified as having type SHORT. But images with more than 64K rows or columns shall use the LONG field type. TIFF readers should accept BYTE, SHORT or LONG values for any unsigned integer field. This allows a single procedure to retrieve any integer value, makes reading more robust and saves disk space in some situations.

Each TIFF field has an associated count. This means that all fields are actually one-dimensional arrays, even though most fields contain only a single value. For example, to store a complicated data structure in a single private field, use the UNDEFINED field type and set the count to the number of bytes required to hold the data structure.

5.1.3. Vendor unique information

Each software manufacturer may choose to store additional information in a standard ISO TIFF in the form of private tags or private tag- values. This can be done by e.g. obtaining private tags and/or tag-values for TIFF 6.0 tags from the Adobe Developers Desk. In a TIFF/A no private tags are allowed, because of the possible lack of documentation.

5.2. Image Data

The image width, i.e. horizontal or X dimension, is recorded as a binary value in the ImageWidth tag-field. The image width may be the shorter or longer dimension of the image, depending upon the orientation of the camera during image capture. The image orientation is defined by the Orientation tag-field. The image length, i.e. vertical or Y dimension, is recorded as a binary value in the ImageLength tag-field. The camera's desired image output rendering resolution in the X-dimension, i.e. the horizontal dimension when the camera is normally oriented, is recorded using the XResolution tag-field, while the output resolution in the Y-dimension is recorded in the YResolution tag-field. The ResolutionUnits tag-value gives the units for the XResolution and YResolution values. These mandatory TIFF 6.0 tag-fields are typically used to determine the default size of the image on the screen. The only valid TIFF/A pixel aspect ratio is 1:1 (square), so that XResolution equals YResolution.

The number of color components or samples per pixel in the image is recorded using the SamplesPerPixel tag- field as a binary value. For example, an image captured using a monochrome sensor has only 1 color component or sample per pixel, while a 3-sensor color RGB camera has 3 color components or samples per pixel. The number of bits needed to store each of the color components (samples) is recorded using the BitsPerSample tag- field as a set of binary values. In the case of a monochrome image, the BitsPerSample tag-field contains only one value, equal to the actual number of bits per pixel.

In the case of an RGB image having 3 color samples per pixel, the BitsPerSample tag-field contains 3 values equal to the actual number of bits of storage used to store each component or sample. In the later case, the number of bits for each color-component could be different, and hence is explicitly stated. The image data is stored using either strips or tiles, which are collectively termed segments. If strips are used, the following tag-fields define the number of strips and the number of rows of image data stored in each strip: StripOffsets, RowsPerStrip, and StripByteCounts. The image shall be divided into an integral number of strips, from 1 strip to the maximum number of strips, which equals the image's length. If necessary, the final strip can be "padded" with zeros.

TIFF/EP recommends that the image data, prior to compression, not exceed 64 Kbytes per strip. This value is chosen to maximize compatibility with various operating systems. The StripOffsets tag-field stores the offsets from the start of the image file to the start of each image data strip. In this way, the reader can easily access various parts of the image. The number of rows per strip are stored in the RowsPerStrip tag-field. The number of image data bytes stored within each strip are recorded in the StripByteCounts tag-field. The "strip" mechanism is very useful in accessing images, because it uses less buffer memory than would otherwise be needed to read in the entire image, all at one time. The order of the image strips is from the top to the bottom of the image.

Example:
ImageWidth = 768
ImageLength = 512
PhotometricInterpretation = 2 (RGB)
SamplesPerPixel = 3
BitsPerSample = 8,8,8
PlanarConfiguration = 1 (Chunky), i.e. RGBRGBRGB..

The size in bytes of each row in this image is 768 PixelsPerRow * (3 SamplesPerPixel * 8 BitsPerSample)/8 BitsPerByte ==> 2304 BytesPerRow. Assuming 8 rows of image data in each strip, the number of bytes per strip is 8 RowsPerStrip * 2304 BytesPerRow ==> 18,432 BytesPerStrip. The number of strips equals 512 Rows/8 RowsPerStrip ==> 64 strips.

5.3. Thumbnail images

There may be more than one IFD in a TIFF 6.0 file. Each IFD defines a subfile. One potential use of subfiles in TIFF 6.0 is to describe related images, such as the pages of a facsimile transmission. A Baseline TIFF 6.0 reader is not required to read any IFDs beyond the first one. In TIFF/A files, the 0th IFD must be an image that can be read by a baseline TIFF 6.0 reader. The thumbnail of the first entry must not be compressed, and must be stored in strips, in order to be fully compatible with TIFF 6.0. The full-resolution image is stored uncompressed as a baseline-readable TIFF image, the full-resolution image could be stored in the 0th IFD. However, TIFF/A recommends that a thumbnail image be stored in the 0th IFD. This provides a version of the image that is small (relative to the full-resolution image) and that may be quickly accessed by reader software. TIFF/A requires that the thumbnail image be stored in strips, so that the thumbnails may be read by any baseline TIFF 6.0 reader. The dimensions of the thumbnail image are restricted by TIFF/EP to 256 pixels maximum horizontally and 256 pixels maximum vertically. The following tag-fields are necessary to define the number of strips and the number of rows of thumbnail image data stored in each strip: StripOffsets, RowsPerStrip, and StripByteCounts.
In the example below, the thumbnail image has one eighth the number of lines and one eighth the number of pixels per line as its parent image. The thumbnail image is a single strip which holds the thumbnail image data. The StripOffsets tag-field stores the offset from the start of the image file to the start of the thumbnail image data strip.
The number of rows per strip, i.e. 64 rows, is stored in the RowsPerStrip tag-field. The number of thumbnail image data bytes stored in the strip is recorded in the StripByteCounts tag-field.
Assumptions:

- Parent ImageWidth = 768
- Parent ImageLength = 512
- Parent PhotometricInterpretation = 2 (RGB)

- Parent SamplesPerPixel = 3
- Parent BitsPerSample = 8,8,8
- Parent PlanarConfiguration = 1 (Chunky), i.e. RGBRGBRGB...
- Thumbnail ImageWidth = 96
- Thumbnail ImageLength = 64
- Thumbnail PhotometricInterpretation = 2 (RGB)
- Thumbnail SamplesPerPixel = 3
- Thumbnail BitsPerSample = 8,8,8
- Thumbnail PlanarConfiguration = 1 (Chunky), i.e. RGBRGBRGB...

The size in bytes of the single thumbnail image strip is 96 PixelsPerRow * (3 SamplesPerPixel * 8 BitsPerSample)/8 BitsPerByte ==> 288 BytesPerRow.
We are storing 64 rows of thumbnail image data in the strip, hence the number of bytes per strip is 64 RowsPerStrip * 288 BytesPerRow ==>18,432 BytesPerStrip.
The number of strips is equal to 64 Rows/64 RowsPerStrip ==> 1 strip.

6. TIFF/A tag definition

6.1. TIFF/A tag list

Tag	Name	Description	TIFF/A	Rating	Comment
262	PhotometricInterpretation	0 = White is zero 1 = Black is zero 2 = RGB	mandatory allowed values: 1 and 2		An image without this tag can not be interpreted at all. Value 0 is not common and it has no advantages.
259	Compression	Compression scheme for data, 1 = no compression	mandatory only value 1 allowed		Image compression is not allowed for archival purposes
257	ImageLength	Number of vertical scan lines	mandatory		An image without this tag can not be interpreted at all.
256	ImageWidth	Number of horizontal pixels (pixels per line)	mandatory		An image without this tag can not be interpreted at all.
296	ResolutionUnit	Unit of physical dimensions of original image	optional	Red	This option is important if no scale is visible in the scene and the

					dimensions of the original are not known.
282	XResolution	Horizontal dimension	optional	Red	This information is a derivative of tag 296 and 257, therefor it should be included.
283	YResolution	Vertical Dimension	optional	Red	
258	BitsPerSample	Number of bits/sample	required allowed values: 8, 16		This information is a derivative of tag 296 and 256, therefor it should be included.
278	RowsPerStrip	See TIFF specification	mandatory		An image without this tag can not be interpreted.
273	StripOffsets	See TIFF specification	mandatory		An image without this tag can not be interpreted.
279	StripByteCounts	See TIFF specification	mandatory		An image without this tag can not be interpreted.
277	SamplesPerPixel	Number of samples / channels per pixel. Gray value images to have 1, color images 3 samples per pixel	mandatory allowed values 1 or 3		An image without this tag can not be interpreted.
315	Artist	Artist or DC Creator	optional	Red	This metainformation is very a standard field in almost every metadata scheme. It should be filled with content.
265	CellLength	Used for dithering	forbidden		Dithering is a not allowed concept of tonal representation in an image.
320	ColorMap	Colormap for palette color images	forbidden Palette color images not allowed		palette colors are not supported for archival purposes.
33432	Copyright	Copyright notice	optional	Orange	Digital rights are important and if possible this tag

					should be filled. Unfortunately in some cases the rights situation is not clear therefore at least a comment should be inserted.
306	DateTime	Date and Time of digital image creation Format: YYYY:MM:DD HH:MM:SS	optional	Green	The date of image creation doesn't specify which event is meant: The date of the creation of the original or the date of creation of the scan.
388	ExtraSamples	Additional components for image like masking etc.	forbidden		Additional layers are not allowed in a TIFF/A
266	FillOrder	See TIFF specification	forbidden		This option is not allowed in a TIFF/A
289	FreeByteCount	See TIFF specification	forbidden		This option is not allowed in a TIFF/A
291	GrayResponseCurve	Optical density for each pixel value	optional for gray value image, else forbidden	Green	Optical densities are only necessary for special applications.
290	GrayResponseUnit	Precision of optical density values	mandatory, if GrayResponseCurve is used, else forbidden	Green	Optical densities are only necessary for special applications.
316	HostComputer	Name of computer where image has been created	optional	Green	The computer name only is only in rare situations important.
270	ImageDescription	Short description of image content	optional (recommended)	Red	A brief embedded information about the image content is important and can be seen as an increasing factor for redundancy.
271	Make	Scanner Manufacturer	optional	Green	The scanner name only is in rare situations

					important, eg for hardware calibration.
281	MaxSampleValue	Maximum component value	forbidden		
281	MinSampleValue	Minimum component value	forbidden		
272	Model	Scanner model	optional	Green	The scanner model only is in rare situations important, eg for hardware calibration.
254	NewSubfileTye	See TIFF specification	forbidden		
274	Orientation	The orientation of the image with respect to the rows and columns	mandatory value allowed: 1		Orientation must be top-left
284	PlanarConfiguration	How the components of each pixel are stored	mandatory value allowed: 1		The inner structure of the data is given and defined by value 1
305	Software	Software which created the digital image	optional	Green	The software name only is in rare situations important.
255	SubfileType	See TIFF specification	forbidden		
263	Thresholding	See TIFF specification	forbidden		
269	DocumentName	The name of the document from which this image was scanned	optional	Green	This integrated meta information is in most cases not corresponding with other metadata structures. If possible fill it in here.
285	PageName	The name of the page from which this image was scanned	optional	Green	This integrated meta information is in most cases not corresponding with other metadata structures. If possible fill it in here.
297	PageNumber	The page number of	optional	Green	This integrated

		the page from which this image was scanned			meta information is in most cases not corresponding with other metadata structures. If possible fill it in here.
286	XPosition	The X offset in ResolutionUnits of the left side of the image, with respect to the left side of the page	optional, if ResolutionUnit, XResolution and YResolution are given, else forbidden	Green	The offset is not important for archival TIFF/A. The TIFF/A represents a final master file and not RAW data.
287	YPosition	The Y offset in ResolutionUnits of the top of the image, with respect to the top of the page	optional, if ResolutionUnit, XResolution and YResolution are given, else forbidden	Green	The offset is not important for archival TIFF/A. The TIFF/A represents a final master file and not RAW data.
317	Predictor	See TIFF specification	forbidden		
322	TileWidth	Tiling	forbidden		Tiling is not allowed for images that shall be archived.
323	TileLength	Tiling	forbidden		Tiling is not allowed for images that shall be archived.
324	TileOffsets	Tiling	forbidden		Tiling is not allowed for images that shall be archived.
325	TileByteCounts	Tiling	forbidden		Tiling is not allowed for images that shall be archived.
332	InkSet	CMYK	forbidden		CMYK is a not allowed color scheme for archiving. CMYK is only important if images are meant to be directly printed. Due to todays color management workflows the only proper working

					colorspace is one of the common RGB spaces.
334	NumberOfInks	CMYK	forbidden		CMYK is a not allowed color scheme for archiving. CMYK is only important if images are meant to be directly printed. Due to todays color management workflows the only proper working colorspace is one of the common RGB spaces.
333	InkNames	CMYK	forbidden		CMYK is a not allowed color scheme for archiving. CMYK is only important if images are meant to be directly printed. Due to todays color management workflows the only proper working colorspace is one of the common RGB spaces.
336	DotRange	CMYK	forbidden		CMYK is a not allowed color scheme for archiving. CMYK is only important if images are meant to be directly printed. Due to todays color management workflows the only proper working colorspace is one of the common RGB spaces.
337	TargetPrinter	CMYK	forbidden		CMYK is a not allowed color scheme for archiving. CMYK is only important if images are meant to be directly

					printed. Due to todays color management workflows the only proper working colorspace is one of the common RGB spaces.
321	HalftoneHints	See TIFF specification	forbidden		Halftone images are not common for archival masters. Grey shades are shown as black spots of different size and not as pixels of grey.
339	SampleFormat	data format of samples	optional, but must be 1 if given (unsigned integer)	Red	This field specifies how to interpret each data sample in a pixel. The most common value is unsigned integer data. Any other format can make the data unreadable
340	SMinSampleValue	See TIFF specification	forbidden		This field specifies the minimum sample value of samples. This tag is not allowed in an image representation where unsigned integers are applied
341	SMaxSampleValue	See TIFF specification	forbidden		This field specifies the maximum sample value of samples. This tag is not allowed in an image representation where unsigned integers are applied
318	WhitePoint	Colorimetry within TIFF	forbidden[1]		This information is typically not used

[1] While these TIFF tags would offer the means of a colormetry for RGB images, these tags have never been used widely and have been replaced by ICC color profiles.

					and has been replaced by ICC profiles.
319	PrimaryChromaticities	Colorimetry within TIFF	forbidden		This information is typically not used and has been replaced by ICC profiles.
301	TransferFunction	Colorimetry within TIFF	forbidden		This information is typically not used and has been replaced by ICC profiles.
342	TransferRange	Colorimetry within TIFF	forbidden		This information is typically not used and has been replaced by ICC profiles.
532	ReferenceBlackWhite	Colorimetry within TIFF	forbidden		This information is typically not used and has been replaced by ICC profiles.
529	YCBCR Coefficients	YCC images	forbidden		The YCC color space is not allowed for image preservation. The only valid color system is RGB in a standard color space like sRGB or Adobe RGB..
530	YCBCR SubSampling	YCC images	forbidden		The YCC color space is not allowed for image preservation. The only valid color system is RGB in a standard color space like sRGB or Adobe RGB..
531	YCBCR Positioning	YCC images	forbidden		The YCC color space is not allowed for image preservation. The only valid color system is RGB in a standard color space like sRGB or Adobe RGB..
512	JPEGProc	See TIFF specification	forbidden		JPEG image data

					compression is not allowed for preservation purposes due to possible compression artefacts.
513	JPEGInterchangeFormat	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
514	JPEGInterchangeFormat Length	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
515	JPEGRestartInterval	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
517	JPEGLosslessPredictors	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
518	JPEGPointTransforms	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
519	JPEGQTables	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible

					compression artefacts.
520	JPEGDCTables	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
521	JPEGACTables	See TIFF specification	forbidden		JPEG image data compression is not allowed for preservation purposes due to possible compression artefacts.
34675	ICC Profile	Extension to TIFF Revision 6. The ICC Profile data is a “dark” object to TIFF that has to conform to the ICC standard	mandatory for color images. The image must be converted either sRGB, AdobeRGB or ProPhoto and the corresponding ICC profile must be stored using this tag.		Color image data stored as RGB by itself doesn’t mean anything. Therefor a clear ICC color space must be applied to give the values of R, G and B a clear physical meaning.
34665	EXIF Tag	Offset if the EXIF IFD	optional[2]	Red	EXIF metadata is commonly used for major metadata. If not used, make sure that another appropriate metadata format is applied.
33723	IPTC Tag	IPTC (International Press Telecommunications Council) metadata.	optional[3] (not recommended)	Red	IPTC metadata is commonly used for major metadata. If not used, make sure that another appropriate metadata format is applied.
700	XMP Tag	eXensible Metadata Platform as proposed	optional (recommended)	Red	Metadata in XMP format is

[2] See below
[3] See below

		by Adobe			commonly used for complex metadata. If not used, make sure that another appropriate metadata format is applied.

6.2. TIFF/A tag definitions group by function

6.2.1. Imagewidth (mandatory)

This tag encodes the number of columns in the image, i.e. the number of pixels per line. This tag is mandatory and there is no default value. The image width may be the shorter or longer dimension of the image, depending upon the orientation of the camera during image capture.
Tag Name = ImageWidth
Tag = 256 (100.H)
Type = SHORT or LONG
N = 1
Value = VALUE (any value representing the image width)
Usage: Main Image, Thumbnail Image

6.2.2. ImageLength (mandatory)

This tag encodes the number of rows in the image, i.e. the number of lines in the image. This tag is mandatory and there is no default value. The image length may be the shorter or longer dimension of the image, depending upon the orientation of the camera during image capture
Tag name = ImageLength
Tag = 257 (101.H)
Type = SHORT or LONG
N = 1
Value = VALUE (any value representing the image length)
Usage: Main Image, Thumbnail Image

6.2.3. NewSubFileType

This tag indicates whether the given IFD is a thumbnail or main (full-size) image. This tag is mandatory and there is no default value. The value shall be explicitly stated. If the

current image is the main image, rather than the "thumbnail" image, a value of 0 shall be encoded in this field.
Tag Name = NewSubFileType
Tag = 254 (FE.H)
Type = LONG
N = 1
Value = VALUE (32 bit field as described below)
Bit 0 1 if the image is a “thumbnail”, else 0
Bit 1 N/A
Bit 2 N/A
Usage: Main Image, Thumbnail Image

6.2.4. SubIFDs

This tag encodes the offsets from the beginning of the file to the location of IFDs that are “treeing” from the current IFD. This tag-value is used to point from the thumbnail IFD to the IFD containing the full resolution image. For a full description refer to the section entitled “Thumbnail Images using ‘SubIFDs’ Trees”.
Tag Name = SubIFDs
Tag = 330 (14A.H)
Type = LONG
N = number of child IFDs
Value = VALUE or VALUE_OFFSET
The Value will contain the offset to the “treed” IFD itself if N=1, otherwise the Value will contain an offset to a location containing an array of offsets to each IFD being “treed” from the current IFD. This array of offsets will contain N entries, i.e. offset pointers to N IFDs. Currently, N=1, and the Value contains the offset to the IFD containing the full resolution image.
Usage: IFD0

6.2.5. Xresolution (optional)

This tag encodes the number of pixels per ResolutionUnit in the ImageWidth direction. This tag specifies the desired ouput rendering Xresolution. This tag is mandatory and there is no default value.
Tag Name = XResolution
Tag = 282 (11A.H)
Type = RATIONAL
N = 1
Value = VALUE_OFFSET (any legal rational number may be used)
Usage: Main Image, Thumbnail Image

(See YResolution, ResolutionUnit)

6.2.6. Yresolution (optional)

This tag encodes the number of pixels per ResolutionUnit in the ImageLength direction. This tag specifies the desired output rendering Yresolution. This tag is mandatory and there is no default value.
Tag Name = YResolution
Tag = 283 (11B.H)
Type = RATIONAL
N = 1
Value = VALUE_OFFSET (any legal rational number may be used)
Usage: Main Image, Thumbnail Image
(See XResolution, ResolutionUnit)

6.2.7. ResolutionUnit (optional)

This tag encodes the unit of measurement for the XResolution and Yresolution. This tag is mandatory and there is no default value.
Tag Name = ResolutionUnit
Tag = 296 (128.H)
Type = SHORT
N = 1
Value = VALUE (possible values listed below)

6.2.8. Orientation (mandatory)

This optional tag encodes the orientation of the camera relative to the scene, when the image was captured. Only the value 1 is allowed.
Tag Name = Orientation
Tag = 274 (112.H)
Type = SHORT
N = 1
Value = VALUE (possible values are listed below)

1	The 0th row represents the visual top of the image, and the 0th column represents the visual left-hand side. The camera orientation is known to be in the normal “landscape” orientation.

3	The 0th row represents the visual bottom of the image, and the 0th column represents the visual right-hand side. Normal orientation rotated 180 degrees
6	The 0th row represents the visual right-hand side of the image, and the 0th column represents the visual top. Normal orientation rotated clockwise 90 degrees.
8	The 0th row represents the visual left-hand side of the image, and the 0th column represents the visual bottom. Normal orientation rotated counter-clockwise 90 degrees.
9	Orientation is unknown.

6.2.9. PhotometricInterpretation (mandatory)

In TIFF/EP, this tag defines the color space of the image data components, and the order of the components. This
tag is mandatory and there is no default value. The value shall be 1,2, or 6 for the thumbnail IFD.
Tag Name = PhotometricInterpretation
Tag = 262 (106.H)
Type = SHORT
N = 1
Value = VALUE (possible values are listed below)

1	BlackIsZero. Used to describe grayscale images. 0 is black, 2**BitsPerSample - 1 is white
2	RGB. Used for images stored in RGB color space. Red, green, and blue minimum intensity is 0, and maximum intensity is 2**BitsPerSample - 1. The component order shall be R, G, B.
6	YCbCr color space. Y refers to the luminance component, Cb and Cr refer to the two chrominance components. The component order shall be Y, Cb, Cr.
32803	CFA. TIFF/EP readers are not required to handle this tag value. Used to describe “raw” image data from single-chip color sensors having a color

	filter array (CFA) overlay. At a given pixel location either a Red, Green, Blue, Cyan, Magenta, Yellow, or White sample value is recorded. For each pixel value, 0 represents minimum intensity, and 2**BitsPerSample - 1 represents maximum intensity. The spatial sampling pattern of the color filter array is defined using the tags CFARepeatPatternDim and CFAPattern. The color of the sample values are defined in the tag CFAPattern as follows; 0=Red, 1=Green, 2=Blue, 3=Cyan, 4=Magenta, 5=Yellow, and 6=White. SamplesPerPixel shall equal 1. The PlanarConfiguration shall equal 1 (Chunky). The component values for each pixel are stored contiguously, as specified by the tags CFARepeatPatternDim and CFAPattern.
Other Values >32767	These values are "Vendor Unique" and are not required to be interpreted by others. These values need to be obtained from Adobe Corporation.

6.2.10. PlanarConfiguration (mandatory)

This tag describes how the components (samples) of each pixel are stored. This tag is mandatory and there is no default value.
Tag Name = PlanarConfiguration
Tag = 284 (11C.H)
Type = SHORT
N = 1
Value = VALUE (possible values listed below)

1	Chunky format. The component values for each pixel are stored contiguously. The order of the components within the pixel is specified by PhotometricInterpretation. For example, chunky data when PhotometricInterpretation equals 2 (RGB) is stored as RGBRGBRGB. The chunky format shall be used when the PhotometricInterpretation value is 1 (grayscale) or 32803 (CFA).
2	Planar format. The components are stored in separate component planes.

6.2.11. SamplesPerPixel (mandatory)

This tag encodes the number of components or samples stored for each pixel in the image. This tag is mandatory and there is no default value. The value shall be explicitly stated.
Tag Name = SamplesPerPixel

Tag = 277 (115.H)
Type = SHORT
N = 1
Value = VALUE (possible values listed below)

6.2.12. BitsPerSample (mandatory)

This tag encodes the number of bits per component for each pixel. This tag is mandatory and there is no default value. This tag provides N values depending upon SamplesPerPixel present.
Tag Name = BitsPerSample
Tag = 258 (102.H)
Type = SHORT
N = SamplesPerPixel
Value = VALUE or VALUE_OFFSET

6.2.13. StripOffsets (mandatory)

This tag encodes the byte offset(s) with respect to the beginning of the file to each strip of the image data. When strips are used, this mandatory tag-field is the only way for a reader to find the image data. TIFF/EP requires that the size of the image data contained in each strip, prior to compression, not exceed 64
KBytes.
Tag Name = StripOffsets
Tag = 273 (111.H)
Type = SHORT or LONG
N = StripsPerImage (when PlanarConfiguration equals 1)
= SamplesPerPixel * StripsPerImage (when PlanarConfiguration equals 2)
Value = VALUE or VALUE_OFFSET
Refer to the tag RowsPerStrip to obtain the derived value for StripsPerImage.
Usage: Main Image, Thumbnail Image
(See PlanarConfiguration, RowsPerStrip)

6.2.14. RowsPerStrip (mandatory)

This tag encodes the number of rows per strip within the image. This tag is mandatory when the image is stored as strips.
Tag Name = RowsPerStrip
Tag = 278 (116.H)
Type = SHORT or LONG

N = 1
Value = VALUE (any value up to and including ImageLength)
Note that the number of StripsPerImage equals floor ((ImageLength + RowsPerStrip - 1) / RowsPerStrip)
Usage: Main Image, Thumbnail Image
(See ImageLength, StripOffsets, StripByteCounts, TileWidth, TileLength, TileOffsets, TileByteCounts)

6.2.15. StripByteCounts (mandatory)

This tag encodes the number of bytes present in each strip after compression has been applied. This tag is mandatory when the image is stored as strips.
Tag Name = StripByteCounts
Tag = 279 (117.H)
Type = SHORT or LONG
N = StripsPerImage (when PlanarConfiguration equals 1)
= SamplesPerPixel * StripsPerImage (when PlanarConfiguration equals 2)
Value = VALUE or VALUE_OFFSET
Refer to the tag RowsPerStrip to obtain the derived value for StripsPerImage.

6.2.16. Compression (mandatory)

This tag encodes the compression scheme used to store the image data. This tag is mandatory and there is no default value. The tag value shall equal “1” for the thumbnail IFD.
Tag Name = Compression
Tag = 259 (103.H)
Type = SHORT
N = 1
Value = 1

6.2.17. ImageDescription (optional, recommended)

A string that describes the subject or purpose of the image is recorded as a null-terminated ASCII string. This tagfield may be additionally used to provide any other type of information related to the image. This is a mandatory tag, with an allowed value of “ ” (null character), indicating “unknown”.
Tag Name = ImageDescription
Tag = 270 (10E.H)
Type = ASCII
N = any (this count includes the null terminating byte at the end of the string)

Value = VALUE or VALUE_OFFSET
Usage: IFD0

6.2.18. Artist (optional, recommended)

This tag encodes the name of the camera owner or image creator.
Tag Name = Artist
Tag = 315 (13B.H)
Type = ASCII
N = any (this count includes the null terminating byte at the end of the string)
Value = VALUE or VALUE_OFFSET
Usage: IFD0

6.2.19. Copyright (optional, recommended)

This tag encodes the copyright notice of the copyright holder. This is a mandatory tag, with an allowed value of " " (null character), indicating "unknown". The complete copyright statement should be listed in this field including any dates and statements of claims. If desired, this tag-field can also list the royalty clearance house.
Tag Name = Copyright
Tag = 33432 (8298.H)
Type = ASCII
N = any (this count includes the null terminating byte)
Value = VALUE or VALUE_OFFSET
Usage: IFD0

6.2.20. SecurityClassification (optional)

This optional tag encodes the level of security classification assigned to the image. The tag value can either be a single ASCII character or an ASCII string.
Tag Name = SecurityClassification
Tag = 37394 (9212.H)
Type = ASCII
N = 1 or any
Value = VALUE or VALUE_OFFSET
The allowed single characters are T (= Top Secret), S (= Secret), C (= Confidential), R (= Restricted), U (=Unclassified). These definitions are based on the NITF security classifications defined in the NITF (National Imagery Transmission Format) specification (MIL-STD-2500).
The multi-character ASCII string tag-value can include one or more of the following NITF fields. The NITF fields should appear in the order listed below. Refer to MIL-STD-

2500 for a complete description of these fields. A NULL is inserted between each NITF field and it's associated value. When multiple NITF fields are used, a NULL character is inserted between each NITF field.

6.2.21. DateTimeOriginal (optional, recommended)

This mandatory tag encodes the date and time the original image was photographed. This tag should never be changed after it is written in the camera or image capture device.
Tag Name = DateTimeOriginal
Tag = 36867 (9003.H)
Type = ASCII
N = 20
Value = VALUE_OFFSET (the syntax of the null terminated string at the specified offset is given below) YYYY:MM:DD HH:MM:SS , with hours 0-24, a space character between the date and time, and a null termination byte. If the camera has no clock, or the clock is disabled, the time should be indicated as 0000:00:00 00:00:00
Usage: IFD0

6.2.22. DateTime (optional)

This mandatory tag encodes the date and time the image was last modified.
Tag Name = DateTime
Tag = 306 (132.H)
Type = ASCII
N = 20
Value = VALUE_OFFSET (the syntax of the null terminated string at the specified offset is given below) YYYY:MM:DD HH:MM:SS , with hours 0-24, a space character between the date and time, and a null termination byte. If the camera has no clock, or the clock is disabled, the time should be indicated as 0000:00:00 00:00:00
Usage: IFD0

6.2.23. TimeZoneOffset (optional)

This optional tag encodes the time zone of the camera clock (relative to Greenwich Mean Time) used to create the DataTimeOriginal tag-value when the picture was taken. It may also contain the time zone offset of the clock used to create the DateTime tag-value when the image was modified.
Tag Name = TimeZoneOffset
Tag = 34858 (882A.H)
Type = SSHORT

N = 1 or 2
Value = VALUE
The allowed values are -12 to +11.

6.2.24. ExposureTime (optional)

This optional tag encodes the actual exposure time used when the image was captured. The units are (fractional) seconds. For example, an exposure time of 1/60 second is encoded as 1/60. The exposure time may be specified by using a single number, if the value is exactly known. Alternately, two values may be used to indicate the range of uncertainty in the exposure time setting. In this case, the first value shall be the minimum time and the second shall be the maximum.
Tag Name = ExposureTime
Tag = 33434 (829A.H)
Type = RATIONAL
N = 1 or 2
Value = VALUE_OFFSET (any valid rational number)

6.2.25. ShutterSpeedValue (optional)

This optional tag encodes the shutter speed value (APEX time value) used when capturing the image. The units are APEX (Additive Systems of Photographic Exposure) values.
Tag Name = ShutterSpeedValue
Tag = 37377 (9201.H)
Type = RATIONAL
N = 1
Value = VALUE_OFFSET (any valid rational number)
Usage: IFD0

6.2.26. Fnumber (optional)

This optional tag encodes the actual lens f-number (ratio of lens aperture to focal length) used when the image was captured. The f-number may be specified by using a single number, if the value is exactly known. Alternately, two values may be used to indicate the range of uncertainty in the f-number setting. In this case, the first value shall be the minimum f-number and the second shall be the maximum.
Tag Name = FNumber
Tag = 33437 (829D.H)
Type = RATIONAL
N = 1 or 2

Value = VALUE_OFFSET (any valid rational number)

6.2.27. ApertureValue (optional)

This tag encodes the actual lens aperture (Av) used when capturing the image. The units are APEX. The maximum value is 99.99, the minimum value is 0.0. In APEX units, a value of 0.0 corresponds to f/1.0, and a
value of 1.0 corresponds to f/1.4.
Tag Name = ApertureValue
Tag = 37378 (9202.H)
Type = RATIONAL
N = 1
Value = VALUE_OFFSET (any valid rational number)
Usage: IFD0

6.2.28. MaxApertureValue (optional)

This optional tag encodes the maximum possible aperture opening (minimum lens f-number) of the camera or image capturing device, using APEX units. The allowed range is 0.00 to 99.99.
Tag Name = MaxApertureValue
Tag = 37381 (9205.H)
Type = RATIONAL
N = 1
Value = VALUE_OFFSET
Usage: IFD0

6.2.29. CellLength (forbidden)

The length of the dithering or half toning matrix used to create a dithered or halftoned bilevel file.

Tag = 265 (109.H)
Type = SHORT
N=1
This field should only be present if Threshholding = 2
No default. See also Threshholding.

6.2.30. ColorMap (forbidden)

Color map for palette color images.
Tag = 320 (140.H)

Type = SHORT
N = 3 * (2**BitsPerSample)

This field defines a Red-Green-Blue color map (often called a lookup table) for palette-color images. In a palette-color image, a pixel value is used to index into an RGB lookup table. For example, a palette-color pixel having a value of 0 would be displayed according to the 0th Red, Green, Blue triplet.

In a TIFF ColorMap, all the Red values come first, followed by the Green values, then the Blue values. The number of values for each color is 2**BitsPerSample. Therefore, the ColorMap field for an 8-bit palette-color image would have 3 * 256 values.

The width of each value is 16 bits, as implied by the type of SHORT. 0 represents the minimum intensity, and 65535 represents the maximum intensity. Black is represented by 0,0,0, and white by 65535, 65535, 65535.

See also PhotometricInterpretation—palette color. No default. ColorMap must be included in all palette-color images.

6.2.31. HostComputer (optional)

The computer and/or operating system in use at the time of image creation.
Tag = 316 (13C.H)
Type = ASCII

6.2.32. Make (optional)

The scanner manufacturer.
Tag = 271 (10F.H)
Type = ASCII
Manufacturer of the scanner, video digitizer, or other type of equipment used to generate the image. Synthetic images should not include this field.

6.2.33. Model (optional)

The scanner model name or number.
Tag = 272 (110.H)
Type = ASCII
The model name or number of the scanner, video digitizer, or other type of equip- ment used to generate the image.

6.2.34. Software (optional)

Name and version number of the software package(s) used to create the image.
Tag = 305 (131.H)
Type = ASCII

6.2.35. DocumentName (optional)

The name of the document from which this image was scanned.
Tag = 269 (10D.H)
Type = ASCII

6.2.36. PageName (optional)

The name of the page from which this image was scanned.
Tag = 285 (11D.H)
Type = ASCII
No default.

6.2.37. PageNumber (optional)

The page number of the page from which this image was scanned.
Tag = 297 (129.H)
Type = SHORT N=2
This field is used to specify page numbers of a multiple page (e.g. facsimile) document. PageNumber[0] is the page number; PageNumber[1] is the total number of pages in the document. If PageNumber[1] is 0, the total number of pages in the document is not available.
Pages need not appear in numerical order. The first page is numbered 0 (zero). No default.

6.2.38. XPosition (optional)

X position of the image.
Tag = 286 (11E.H)
Type = RATIONAL
N=1
The X offset in ResolutionUnits of the left side of the image, with respect to the left side of the page.

No default.

6.2.39. YPosition (optional)

Y position of the image.
Tag = 287 (11F.H)
Type = RATIONAL N=1
The Y offset in ResolutionUnits of the top of the image, with respect to the top of the page. In the TIFF coordinate scheme, the positive Y direction is down, so that YPosition is always positive.
No default.

6.2.40. SampleFormat (optional)

Tag = 339 (153.H)
Type = SHORT
N = SamplesPerPixel
This field specifies how to interpret each data sample in a pixel. Possible values are:
1 = unsigned integer data
2 = two's complement signed integer data
3 = IEEE floating point data [IEEE]
4 = undefined data format

Note that the SampleFormat field does not specify the size of data samples; this is still done by the BitsPerSample field.

A field value of "undefined" is a statement by the writer that it did not know how to interpret the data samples; for example, if it were copying an existing image. A reader would typically treat an image with "undefined" data as if the field were not present (i.e. as unsigned integer data).

Default is 1, unsigned integer data.

7. TIFF/A metadata

A data file without metadata can not be reconstructed nor can it's data be interpreted correctly. This includes mainly the technical configuration and contextual information. Therefore metadata is crucial for any application where sustainability is important. Most important is technical metadata, that is mandatory to be able to read a TIF-file without errors. In addition a standard

TIFF offers multiple options for basic metadata storage but the structure and the origin of those are in the photography, press and media area. For example can be found fields for camera settings and capture properties, for copyright entries and a slim description of the content like keywords.

7.1. EXIF

The Exchangeable Image File Format is a widely used metadata standard for photography applications. The aim of this standard is manly to have static camera capture information such as the camera model and information that varies with each image such as aperture, shutter speed, focal length, metering mode, orientation (rotation) or ISO speed information. Besides of that a lot of additional technical meta-information can be stored in this metadata container. In general it is recommended to store as much meta information as possible in a file for archival purposes. It is especially important to enrich the TIFF with contextual metadata that describes the content. Therefore it is recommended to check if all possible EXIF fields are set.

7.2. ITCP

The IPTC metadata contains basically descriptive metadata about image content, creator etc. and copyright information. The layout of the IPTC does not follow the TIFF standard, so it has to be embedded as a binary object within the TIFF file. Therefore it is not recommended to include important descriptive metadata as IPTC tag. Using the native TIFF tags or the XMP-Format is a much better option.

7.3. XMP

XMP is a XML/RDF-based standard for including descriptive metadata into various image file formats. Since its based on XML, it is sort of human readable. XMP is an open-source, public standard that has a widespread support. It is recommended to use XMP metadata because of it’s flexibility. For archival purposes important technical and contextual meta information shall be stored, in addition to including the XMP-header into the TIFF header, in a sidecar file following the XMP-standard. The metadata itself shall be stored following one of the common metadata standards like METS or Dublin core.

8. Annex. (informative) Best practices for TIFF/A

Image capture

Usually, high quality image capture results either in a proprietary raw file, an Adobe DNG-file, or a TIFF/EP. The aim of such a *Digital Negatives* is a as complete set of image data with the possibility to change settings or algorithms for rendering the final file. Each of these files contains numerous camera specific elements and depends on special software to be rendered, a so called RAW converter. A *Digital Negative* can be converted/interpreted with various RAW-converters leading in different results, regarding interpolation quality, color transforms and destination color space. For maximum color fidelity a color profile can be calculated using a standard color chart for camera calibration. that is applied while exporting. Because of the high level of flexibility in interpretation the TIFF/EP and other raw file formats are not suited for preservation.

Digital master for image manipulation

For any modification of the image, the raw file will be typically transformed into a standard TIFF-format. Manipulation can involve cropping, various adjustments of the tonality and the color appearance. In most cases contextual metadata is added. Today TIFF images are still a often chosen variant for image storage for any type of master file. The master files are then transformed into arbitrary files for distribution or any other application.

Archival Master

In an archival environment, where persistence of the format definition is crucial, the TIFF has some major drawbacks. The flexibility of the format and the numerous customizable tags are, as written in the first paragraph, are disadvantageous for data sustainability. Therefore it is necessary to define a subset of allowed, forbidden and mandatory tags and the structure of the format must be clearly defined, following the rules of archiving. This approach is very similar to the limited set of PDF descriptors, leading to PDF/A.
Therefor we recommend in analogy to PDF/A an adjusted TIFF that we call in this context ***TIFF/A (TIFF/ARCHIVE)*** that has been described in detail above.