How to Digitize Historical Pictorial Collections to Preserve Them in a Digital Life Cycle?

Michel PFEIFFER
Swiss Institute for Information Science (SII)
University of Applied Sciences Graubünden
Michel.Pfeiffer@fhgr.ch

Abstract

The creation of collections in archives, libraries, museums, or documentation centres is and remains the linchpin of any practice-oriented information science studies programme. At the Swiss Institute of Information Science, the life cycle of analogue and digital cultural assets was studied early on. With the advent of digital long-term archiving, the need arose to understand and support objects and metadata from the "origin", including during the digitization process. The establishment of a laboratory for the digitization of a wide range of cultural assets, from photography and written material to records and audiovisual media, was therefore a logical conclusion. This Lab made it possible to gain divergent perspectives in the complex process of the different models and life cycles. This paper will discuss the Metamorfoze Digitization standard as well as the technical Exchangeable Image File (Exif) Metadata Format, used for the digital life cycle aka. "digital long-term archive". This article was originally published in German (Pfeiffer, 2020).

Keywords: digital image collections, digitization standard, digital life cycle, Metamorfoze, Exif data

1. Introduction

With the advent of bits and bytes, the cultural technique of collecting and presenting is undergoing a profound upheaval. Today, there is a consensus that memory institutions are no longer solely focused on making content accessible, but also on delivering the content online for a wide variety of applications. This development has been brought about by new, technology-driven services that are competing with the traditional catalogue. Consider platforms such as Flikr, YouTube, and Instagram, or Google's convenient image search. In social media, image content is being "tagged", "liked", recommended, or even commented on like never before. Collecting institutions are therefore facing a huge challenge—technically, financially and, above all, conceptually[1]. These "content" aspects have been increased by the pandemic. Which institution does not ask itself what role it should play in a "dematerialized" cultural society? How can they position themselves with their digital collections on the market?

The following text will not be able to answer these questions. This task falls to each institution individually. The text pursues the fundamental question of what specific requirements digital long-term archiving places on image data so that it cannot only be successfully presented in the long term, but also archived sustainably. This debate does not begin before archiving, when it is usually too late, but long before digitization. This is forcing a paradigm shift, "archive first". The background to this question seems almost trivial: How else can the high, recurring costs for long-term digital archiving or the communication of content be justified?

Anything that wants to be stored digitally for centuries in the future should be able to fulfil well-documented requirements in comprehensible quality. With the International Image Interoperability Framework (IIIF), a (ignored?) technology is already on the horizon that enables image data to be individually compared and used across all platforms (IIIF, 2020). It would be a shame if one's own institution were unable to use this technology in the future or any other—should it become established—simply because the quality of the image stocks is not specific enough to be able to use them in a meaningful way. On the other hand, what justifies the use of taxpayer money to preserve data scrap in OAIS-compliant long-term archiving systems?

The aim of the text is to provide a qualitative basis for ensuring that image data can be archived and presented in the long term. The article describes

[1] This position has already been discussed here: Michel Pfeiffer, How can image collections be evaluated? Selection, preservation and mediation strategies in the context of digitization projects, in: Zeithistorische Forschungen/Studies in Contemporary History (2015) H2, 317-325.

which standards exist for quality assurance, which measurement tools are used for this purpose, and identifies desiderata of the archival ingest process. The author's thesis is that the traditional key properties of digital image data are insufficient for both the object data and the embedded metadata.

2. Digitization Standards

Standardization efforts such as format harmonization are aimed at preventing the file formats to be ingested from growing to apocalyptic proportions. This pleases the Preservation Planning Manager immensely (Library of Congress, 2020). He/she would rather solve four big problems than 1000 small ones. The same applies to the rampant growth of various metadata standards. Two "embedded" standards—IPTC and EXIF[2]—are important for digitization and long-term archiving (IPTC, 2020; EXIF, 2020).

In the digitalization process, standardization efforts are aimed at reducing complexity so that processes can be improved. The documentation of these processes creates comparability and therefore transparency. The basis of all standardization is the desire for comprehensible measurement criteria that allow a statement to be made about the quality of a digitized image. The aim here is not just to obtain "nice looking images", but to acquire image data worthy of archiving, where any loss of information is documented and thus known. Therefore, the focus is less on defining aesthetic characteristics and more on defining technically measurable criteria.

In the last decade, several standards have been developed for the quality assurance of image data. The US standard FADGI should be mentioned here (FADGI, 2016). In cultural institutions, the guidelines initiated in 2004 have been an integral part of the American quality debate since 2007. A "four-star system" describes different tolerances that must be adhered to during digitiza-

[2] Exif is a standard that specifies formats for images, sound, and ancillary tags used by digital cameras incl. smartphones, scanners, and other digitization hardware. Almost all new digital hardware can use the EXIF annotation, storing information on the image such as shutter speed, exposure compensation, F number, what metric system was used, date and time the image was taken, white balance, auxiliary lenses that were used, and resolution. Some images may even store GPS information, which is handy for geographical annotations. It is perhaps not obvious, but this original information will serve as the foundation for future historians. Today, every digital preservation planner will appreciate them.

tion depending on the materials. The current version describes both transparent transmitted light media (negatives and slides) and reflective reflected light media (photography, graphics, paintings, etc.). With Metamorfoze, Hans van Dormolen developed a further guideline on behalf of the Dutch National Museum to ensure that colour differences in particular are minimized in the digitization process of reproductions of paintings (van Dormolen, 2012).[3] The ISO standard TS 19264-1 was recently added and updated by the ISO 19264-1:2021 (ISO/TS 19264-1, 2017; ISO 19264-1, 2021). The first part of the standard, which was adopted in 2017, describes criteria for digitizing photographic, reflective incident light materials. The second part of the standard, which has not yet been published, will deal with transparent media in the future.

The individual standards differ methodically both in their mathematical formulae and in the measurement fields used to assess these criteria, as well as in the respective form of interpretation of the corresponding tolerances in which a quality criterion is deemed acceptable or not. The comparative principle of target and actual values is common to all standardization efforts. Specific measurement fields, so-called targets (Figure 1), provide the target measurement values corresponding to the standard.

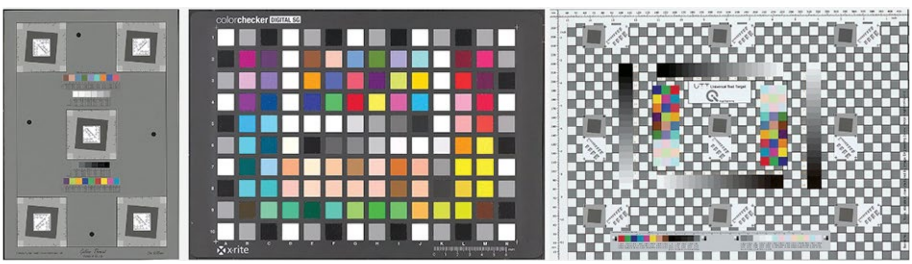

Figure 1. *From left to right: The Golden Thread Object Level Target developed by Don Wiliams for the FADGI guideline, Color Checker Digital Semi Glossy from X-Rite used by Hans van Dormolen for Metamorfoze, and the Universal Test Target (UTT) from Image Engineering for testing the ISO 19264 standard criteria*

The digitization objective is now to reproduce these target values as accurately as possible in practice. The equipment and its configuration play a

[3] Hans van Dormolen has since revised this guideline, but it has not yet been published. See https://www.ingentaconnect.com/contentone/ist/ac/2019/00002019/00000001/art00003 (accessed on 25 May 2020)

decisive role here. Not all devices are able to meet the requirements that the standards place on a digitized image. Hardware and software manufacturers must be measured against this objective. Data that fulfils these very rigid requirements can be reused almost without restriction.

To ensure that the target/actual value comparison is successful, a corresponding measurement field is attached to each original (Figure 2).

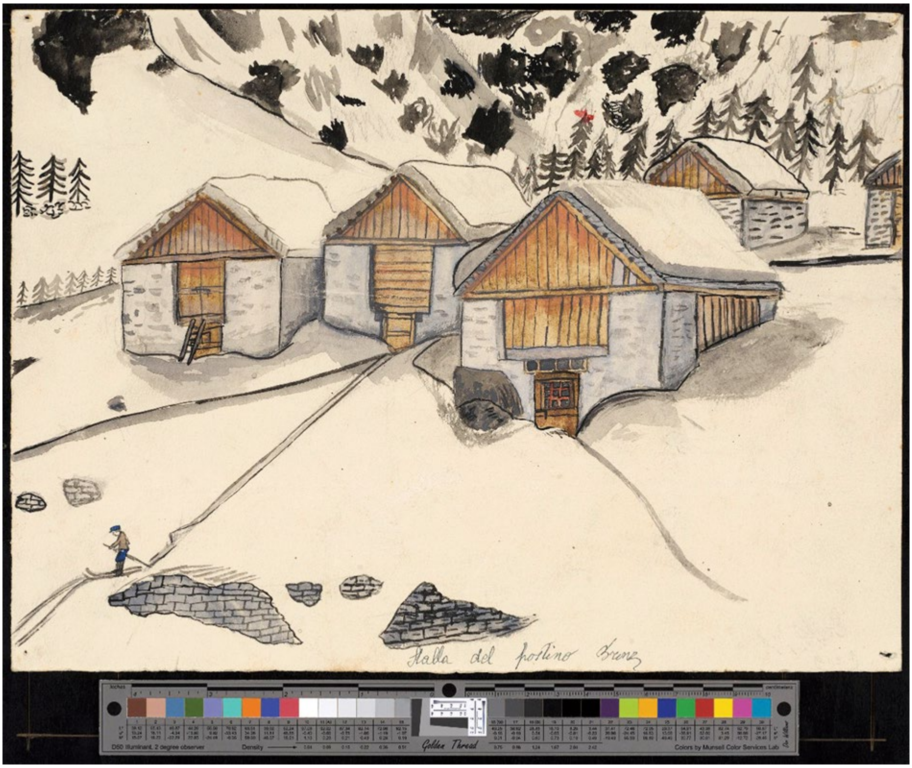

Figure 2. *Signature PKW-009-115, Pestalozzianum Foundation Collection*

Assuming an ideally homogeneous illumination, the measurement target in Figure 2 documents the actual values of the quality criteria to be examined resulting from the digitization. The quality is now determined by comparison. The assessment of quality is always based on an analysis of a "proxy". As such, the resulting quality can never be better than the corresponding measurement field. One must be aware of this fact. By understanding the strengths and weaknesses of the targets used, quality control will be stable over a long period of time and therefore work very well.

The advantage of this approach is obvious. It can be carried out not just once, but throughout the entire future life cycle. The object itself is "inseparably" linked to its actual documentary values—which are also metadata. However, this leads to a larger image data set, which is created by the addition of the measurement field. For very small objects, which are smaller than A5, this can result in major changes to the storage capacity. The requirement quickly increases by dual figure percentage of the original file size. For larger objects between A4 and A2, the narrow strip is less significant; for formats larger than A2, the additional file size is almost negligible.

Production processes can thus be fully documented and checked. If a measured variable is not correct, the production process is checked for the corresponding defects. The strengths and weaknesses of individual scanners or digital cameras can thus be quickly recognized. The same applies to checking the software components used. A coordinated, continuous quality assurance process thus provides production security.

The supplier uses the actual values to prove that his/her end-product fulfils the ordered quality requirement. On the other hand, the purchaser of a file that must comply with a specific standard will be able to check this immediately and without any doubt in the future. Unpleasant discussions about reworking or price reductions are therefore obsolete.

3. Quality Criteria

While the US-American FADGI standard and Metamorfoze recognize up to thirteen different measurable criteria—alongside Reference File Formats, Colour Space, and Bit Depth—such as White Balance and Tonal Capture, Exposure, Gain Modulation, Noise, Illumination, Colour Cast and Colour Accuracy, Sampling Rate(s), Theoretical Resolution and Real-World Sampling Efficiency, Sharpness, Colour Misregistration, Geometric Distortion, as well as Artifacts such as Newton Rings, Light Reflections, or Dirt and Dust, this article is limited to a few exemplary criteria to simplify the introduction to the topic.

If an institution places an order according to one of the three Metamofoze guidelines (full, light, or extralight), the standard already defines the permitted file format, the necessary colour depth, and the working colour space (Dormolen, 2012). These three simple influencing factors can still be easily checked using familiar methods such as Jhove and Exif Toolkit (Jhove, 2020; Exif Toolkit, 2020). Depending on the original format, Metamorfoze also determines the required resolution (Dormolen, 2012). This trivial criterion—it only determines how many pixels per inch (ppi) are read out—becomes complex when the resolution is considered in terms of sharpness. This is because the measure of 600 pixels per inch, for example, provides no indication about the sharpness of these pixels. High-resolution but blurred images manifest neither short-term nor long-term meaningfulness. In technical jargon, this criterion is called resolution efficiency. The sharpness of an image motif is therefore a criterion that must always be assessed in relation to contrast and resolution. The best possible contrast in turn requires correct exposure. This outlines the interaction between light, exposure, resolution, and the resulting sharpness and colour accuracy. Colour fidelity is another critical characteristic. Here, the colour difference DeltaE[4] between the target value of the target and the actual value achieved in the digitized image is compared. Metamorfoze recognizes possible tolerances. A data set may well have a single statistical outlier between DeltaE 18-10 if the sum of all average colour distances does not exceed the value of 5 or 4 for Metamorfoze “full”. This value thus describes the limit of perceptible colour fidelity. Generally speaking, a DeltaE values of 6 and higher are considered to be large colour differences that are immediately perceptible and therefore cannot be tolerated. Values below 1 are almost imperceptible.

The standards presented recognize additional criteria such as the distortion that describes the geometric representation tolerance. In maps and plans, for example, it ensures that a straight line remains a straight line. This enables the use of historical data in digital precise geographic information systems, for example in archaeology. The requirement for true-to-scale digitization is therefore also obvious. The outlined criteria are rounded off by further tech-

[4] In colour science, colour difference is the distance between two colours. Common definitions make use of the Euclidean distance (Delta E) in a device-independent colour space. This metric allows quantified examination of a notion that formerly could only be described with adjectives. Quantification of these properties is of great importance to those whose work is colour critical.

nical and practical factors, such as signal noise, white balance, or tonal gradation. All criteria are operationalized, measurable, and thus objective.[5] There is no “right” standard, nor are there good or bad targets. Ultimately, it is the combination of objectives, equipment, and measuring fields in combination with expertise and experience that ensures quality.

Misuse of the technology cannot be ruled out. Here, “ideal measuring fields” are “copied” into non-ideal images by means of image processing. This may happen in individual cases, but they can be identified very easily. With a little experience, a quality manager will immediately recognize that the values within specific device samples are constantly repeated. However, production processes are subject to changing conditions. These are based on voltage fluctuations in light sources or image sensors, rounding processes in the image-processing software or changes in climatic conditions. In this respect, specialized automatic evaluation systems are necessary.

4. Software Tools

To measure actual values, appropriate tools are needed as well. For example, some values can be analysed manually using Photoshop and Excel. This exercise is certainly useful to understand the theory but is not effective in practice. Unfortunately, the easily accessible platform delt.ae was closed on December 31, 2023. In recent years, this measurement platform has provided low-threshold access to countless learners (Picturae Nederland, 2020). The platform was free to use. Basically, image data was uploaded to a server, analysed, and evaluated. The user could choose between the Metamorfoze and FADGI standards; the ISO standard was not supported.

Based on the appropriate criteria and tolerances, delt.ae displayed the results as “red” or “green”. With another mouse click, the measured values can be analysed individually and in detail. The measurement data can

[5] The respective formulas of the various authors or the interests of the manufacturers involved that led to the standards or targets could be the subject of an excellent debate. However, this discourse is not relevant for direct quality assurance.

also be exported as a CSV file, offering a simple and intuitive approach to the topic, which is nowadays history. Soon, high sophisticated tools such as iQ-Analyzer from Image Engineering or the Golden Thread Analysis Software introduced by Don Williams will take over (iQ-Analyzer, 2020; Image Science Associates, 2020). These solutions are very powerful but expensive for smaller projects and complex for educational purposes. The Golden Thread Analysis Software is very suitable for checking the FADGI standard. It should be mentioned here that the LOC has licensed this procedure, known as DICE, and made it available as an open-source version (FADGI, OpenDICE, and AutoSFR, 2020). Open DICE can be used well in individual cases and for teaching purposes, but larger service projects require faster and more stable tools. This is the domain of iQ-Analyzer from Image Engineering, which can also be used for the ISO quality assurance standard.

5. Metadata

It is important to note here that image data has been checked according to Metamorfoze or another standard as part of a reliable production process, i.e., file formats have also been previously validated as such. As a result of digital production processes, further technical metadata are generated, which flows into the EXIF data. In addition, it is possible to add descriptive information to the IPTC data during the production process. These datasets are significant to the lifecycle.

A central point for the ingest process are the significant properties of the digitized image objects.

"Due to the need for exact definitions of the properties to be preserved ("Significant Properties") (and automatically those aspects that can be neglected or lost) as well as the requirements for the long-term archiving process itself, the creation of the criteria tree ("Objective Tree") provides an enormous gain in understanding. This often makes it clear and obvious for the first time what digital long-term archiving means as a whole. The user must (and thereby will) develop an understanding of the specific properties of the inventory to be archived in order to be able to make correct requirements and decisions."[6] (Neuroth et al., 2020)

If we now examine the depths of the embedded metadata (Table 1), it becomes clear that the differentiation between relevant and less relevant properties requires a well-founded analysis. The following metadata report was created with the Exiftool. It documents the "production history", a central factor in future life cycle archive strategies.

Following Metadata tags of Figure 2; Signature PKW-009-115, Pestalozzianum Foundation Collection:

Table 1. *Metadata report of the signature PKW-009-115 from Figure 2*

Existing Exiftool tags			
ID	**Group**	**Exiftool Tag**	**Embedded value**
1	file	filesize	58 MB
2	file	filemodifydate (w)	2020:05:23 19:56:41+02:00
3	file	fileaccessdate	- 2020:05:23 19:56:41+02:00
4	file	fileinodechangedate	- 2020:05:23 19:56:41+02:00
5	file	filetype	TIFF
6	file	mimetype	image/tiff
7	file	exifbyteorder (w)	Little-endian (Intel, II)
8	file	currentiptcdigest	- ddc955ff1d862aff81a2cc680acbc3de
9	exif	subfiletype (w)	Full-resolution image
10	exif	imagewidth (w)	4919
11	exif	imageheight (w)	4091
12	exif	bitspersample (w)	8 8 8
13	exif	compression (f)	Uncompressed
14	exif	photometricinterpretation (w)	RGB
15	exif	make (w)	Phase One
16	exif	model (f)	P40+

[6] Translated by the author. Original Copy: "Durch den Zwang zu exakten Definitionen zu den zu bewahrenden Eigenschaften ('Significant Properties') (und damit auch automatisch jener Aspekte, die vernachlässigt werden können bzw. verloren gehen dürfen) sowie der Anforderungen an den Langzeitarchivierungsprozess selbst bietet die Erstellung des Kriterienbaumes ('Objective tree') einen enormen Verständnisgewinn. Hierbei wird häufig erstmals bewusst und offensichtlich, was digitale Langzeitarchivierung insgesamt bedeutet. Der Anwender muss (und wird dadurch) ein Verständnis für die spezifischen Eigenschaften des zu archivierenden Bestandes entwickeln, um richtige Anforderungen und Entscheidungen treffen zu können."

Existing Exiftool tags			
ID	**Group**	**Exiftool Tag**	**Embedded value**
17	exif	stripoffsets	-28924
18	exif	orientation (f)	Horizontal (normal)
19	exif	samplesperpixel (w)	3
20	exif	rowsperstrip (w)	4091
21	exif	stripbytecounts	60370887
22	exif	xresolution (w)	400
23	exif	yresolution (w)	400
24	exif	planar configuration (w)	Chunky
25	exif	resolutionunit (w)	inches
26	exif	software (f)	Adobe Photoshop CC 2014 (Windows)
27	exif	modifydate (w)	2018:12:18 14:06:50
28	exif	artist (f)	Pestalozzianum Foundation
29	exif	copyright (w)	Pestalozzianum Foundation
30	exif	exposuretime (f)	1/100
31	exif	fnumber (w)	11
32	exif	exposureprogramme (w)	Manual
33	exif	iso (w)	100
34	exif	exifversion (w)	230
35	exif	datetimeoriginal (w)	2017:07:25 10:38:53
36	exif	createdate (f)	2017:07:25 10:38:53
37	exif	shutterspeedvalue (w)	1/100
38	exif	aperturevalue (w)	11
39	exif	exposurecompensation (f)	0
40	exif	metreingmode (f)	Centre-weighted average
41	exif	lightsource (f)	Other
42	exif	focallength (w)	80.0 mm
43	exif	colourspace (w)	Uncalibrated
44	exif	exifimagewidth (w)	4919
45	exif	exifimageheight (w)	4091
46	exif	focalplanexresolution (w)	1666.666656
47	exif	focalplaneyresolution (w)	1666.666656
48	exif	focalplaneresolutionunit (w)	cm

Existing Exiftool tags			
ID	**Group**	**Exiftool Tag**	**Embedded value**
49	exif	filesource (f)	Digital Camera
50	exif	scenetype (w)	Directly photographed
51	exif	whitebalance (w)	Unknown (5)
52	exif	imageuniqueid (f)	00E058000065000C04010C1E8E008BB1
53	exif	serialnumber (w)	EG031678
54	exif	lensinfo (w)	1mm f/2.799996445-21.99997864
55	exif	lensmodel (f)	Schneider Kreuznach LS 80mm f/2.8
56	xmp	xmptoolkit (w)	Adobe XMP Core 5.5-c021 79.155772, 2014/01/13-19:44:00
57	xmp	creatortool (w)	Capture One Pro 10.1.2 Windows
58	xmp	metadatadate (w)	2018:12:18 14:06:50+05:30
59	xmp	lens (f)	Schneider Kreuznach LS 80mm f/2.8
60	xmp	imagenumber (w)	35761
61	xmp	firmware (w)	P40+-M, Firmware: Main=5.2.2, Boot=2.3, FPGA=1.2.4, CPLD=5.0.1, PAVR=1.0.3, UIFC=1.0.1, TGEN=1.0
62	xmp	legacyiptcdigest (w)	3D557870844B48BA8DF0F04C0774AAC2
63	xmp	colormode (w)	RGB
64	xmp	iccprofilename (w)	eciRGB v2
65	xmp	creatorpostalcode (w)	8090
66	xmp	creatoraddress (w)	Lagerstrasse 2
67	xmp	creatorcity (w)	Zurich
68	xmp	creatorcountry (w)	Switzerland
69	xmp	rights (f)	Pestalozzianum Foundation
70	xmp	creator (f)	Pestalozzianum Foundation
71	xmp	format (w)	image/tiff
72	xmp	instanceid (f)	xmp.iid:74081619-a391-374c-871d-51229058e9b8
73	xmp	documentid (f)	adobe:docid:photoshop:107d76a7-02a0-11e9-8df7-f29931c0638f
74	xmp	originaldocumentid (w)	xmp.did:9ffaf707-21d3-cc48-8ca8-3de-a956b705f
75	xmp	historyaction	saved, saved
76	xmp	historyinstanceid	xmp.iid:9ffaf707-21d3-cc48-8ca8-3dea956b705f, xmp.iid:74081619-a391-374c-871d-51229058e9b8

Existing Exiftool tags			
ID	**Group**	**Exiftool Tag**	**Embedded value**
77	xmp	historywhen	2018:12:18 14:06:50+05:30, 2018:12:18 14:06:50+05:30
78	xmp	historysoftwareagent	Adobe Photoshop CC 2014 (Windows), Adobe Photoshop CC 2014 (Windows)
79	xmp	historychanged	/, /
80	iptc	codedcharacterset (w)	UTF8
81	iptc	applicationrecordversion (w)	4
82	iptc	by-line (w)	Pestalozzianum Foundation
83	iptc	source (w)	Archive children & youth drawing
84	iptc	datecreated (w)	2017:07:25
85	iptc	timecreated (w)	10:38:53+00:00
86	iptc	digitalcreationdate (f)	2017:07:25
87	iptc	supplementalcategories (w)	new supplemental category
88	iptc	copyrightnotice (w)	Pestalozzianum Foundation
89	photoshop	iptcdigest (w)	ddc955ff1d862aff81a2cc680acbc3de
90	photoshop	displayedunitsx (w)	inches
91	photoshop	displayedunitsy (f)	inches
92	photoshop	globalangle (f)	30
93	photoshop	globalaltitude (f)	30
94	photoshop	photoshopthumbnail	(Binary data 7896 bytes, use -b option to extract)
95	icc_profile	profilecmmtype	ADBE
96	icc_profile	profile version	02.04.2000
97	icc_profile	profileclass	Display Device Profile
98	icc_profile	colourspacedata	RGB
99	icc_profile	profileconnectionspace	XYZ
100	icc_profile	profiledatetime	2007:03:02 10:07:41
101	icc_profile	profilefilesignature	acsp
102	icc_profile	primaryplatform	Unknown ()
103	icc_profile	cmmflags	Not Embedded, Independent
104	icc_profile	devicemanufacturer	
105	icc_profile	devicemodel	
106	icc_profile	deviceattributes	Reflective, Glossy, Positive, Colour
107	icc_profile	renderingintent	Perceptual

Existing Exiftool tags			
ID	**Group**	**Exiftool Tag**	**Embedded value**
108	icc_profile	connectionspaceilluminant	0.9642 1 0.82491
109	icc_profile	profilecreator	bICC
110	icc_profile	profileid	9c6d34a5ada445f6146d98b0510c126d
111	icc_profile	profilecopyright (w)	Copyright (C) 2007 by Color Solutions, All Rights Reserved. Licence details can be found on: http://www.eci.org/eci/en/eciRGB.php
112	icc_profile	profiledescription	eciRGB v2
113	icc_profile	mediawhitepoint	0.9642 1 0.82491
114	icc_profile	redtrc	(Binary data 1412 bytes, use -b option to extract)
115	icc_profile	greentrc	(Binary data 1412 bytes, use -b option to extract)
116	icc_profile	bluetrc	(Binary data 1412 bytes, use -b option to extract)
117	icc_profile	redmatrixcolumn	0.65027 0.32028 0
118	icc_profile	greenmatrixcolumn	0.17804 0.60205 0.06783
119	icc_profile	bluematrixcolumn	0.13588 0.07767 0.75708
120	composite	aperture	11
121	composite	datetimecreated	2017:07:25 10:38:53+00:00
122	composite	imagesize (w)	4919x4091
123	composite	scalefactor35efl	1.1
124	composite	shutterspeed	1/100
125	composite	circleofconfusion	0.027 mm
126	composite	fov	22.6 deg
127	composite	focallength35efl	80.0 mm (35 mm equivalent: 90.2 mm)
128	composite	hyperfocaldistance	21.83 m
129	composite	lightvalue	13.6

This metadata report only includes data that is embedded in the file. Classic descriptive catalogue metadata is not yet included here. Let us derive the narrative from it.

The file PKW-009.115 was digitized manually on 25 June 2017 at 10:35 am (ID 35, 36 & 84, 85) with a Phase One camera (ID 15) of type P 40 + (ID 16) on which an 80 mm (ID 42, 127) Schneider Kreuznach lens of quality class LS (ID 55) was mounted, with a shutter speed of 1/100 second (ID 37, 124) and aperture 11 (ID 38) at a sensitivity of 100 ISO (ID 33).

The camera can be unequivocally identified by its serial number (ID 53). The firmware version 5.2.2 (ID 61) installed in the camera is also visible. The RAW data of this camera was "developed" with the Capure One software version 10.1.2 (ID 57) on a Windows computer as an uncompressed tiff (ID 5, 13) with a colour depth of 8 bits per channel (ID 12) in the RGB colour space (ID 14), to which the ECI-RGB v2 colour profile (ID 64) is attached. The central colorimetric specifications are also documented. This includes the date of the ECI-RGBv2 profile (ID 100), its XYZ orientation (ID 90) with the white point definition (ID 108 and 113), and the respective channel orientation of the R, G, and B channels (ID 117-119). The image is 58 MB in size (ID 1).

The landscape-format file (ID 18) with the dimensions 4919 pixels x 4091 pixels (ID 10,11) and a resolution of 400ppi (ID 22, 23) was opened and saved on 18 December 2018 at 14:06 (ID 27) with a Windows operating system on which a Photoshop version CC from 2014 (ID 26,78) was installed. The data says nothing about the editing, only that Photoshop was configured in inches (ID 90/91). The image could also just have been opened for visual inspection and saved incorrectly. The real size can also be calculated from the technical data. A width of 4919 pixels at 400 ppi corresponds to 12.3 inches, which is 31.2 cm. The height is 26 cm. The classic recording of the object size thus becomes almost obsolete with true-to-scale digitization. At the time of production, the ownership of the file is taken from the descriptive IPTC data: Pestalozzianum Foundation, Lagerstrasse 2, 8090 Zurich, Switzerland. The source reference (ID83) refers to the holdings from the archive of children's and youth drawings.

Any photo historian would be delighted to have this information from historical collections from the early days of photographic history. Similarly, every preservation manager will find this information valuable in order to assess a migration that might force him/her to convert image data correctly in terms of colourimetry (Table 2). The data from Table 1 and Table 2 will also provide future media archaeologists with a starting point for possible reconstruction attempts.

Table 2. *Measurement results of the Metamorfoze verification of the signature PKW-009-115 (Figure 2). The results shown were created using the delt.ae platform and downloaded as a CSV*

Patch	R	G	B	L	a	b2
1	118.5	95.2	82.8	40.29	12.96	12.72
2	190	161.7	147.1	66.92	15.79	15.06
3	115.2	132.4	166.1	51.2	-5.14	-22.45
4	108	125.4	79.6	46.02	-15.29	22.05
5	142.4	138.4	185.1	56.54	8.05	-25.53
6	149.5	202	191.6	73.28	-32.45	-3.8
7	209.1	142.3	71.4	64.98	36.4	57.38
8	94.1	94	175.6	40.85	12.48	-45.17
9	179.5	102.9	113.4	53.21	43.61	15.6
10	255	254.7	255	99.92	0.23	-0.12
11	243.4	242.9	248.3	95.49	0.91	-2.97
12	229.6	229.5	233.2	90.14	0.46	-2.08
13	216.6	216.6	219.4	85.04	0.31	-1.57
14	189.6	189.9	192.5	74.52	0.06	-1.52
15	163.3	163.7	165.7	64.21	-0.03	-1.24
16	130.4	129.7	131.9	51.02	0.69	-1.11
17	103.1	103	103.9	40.44	0.13	-0.45
18	76.6	77.6	77.4	30.29	-0.57	-0.09
19	50.1	49.4	50.8	19.5	0.52	-0.62
20	30.1	28.1	29.1	11.31	1.24	-0.14
21	22.6	18.9	21.2	7.96	2.4	-0.49
22	93.7	71.4	112.9	32.67	16.99	-18.26
23	180.9	202.9	94.6	74.76	-22.58	54.14
24	218.9	177.9	67.9	73.8	19.63	69.09
25	63	68.8	151.7	30.84	11.15	-46.41
26	114.5	165	98.9	57.76	-38.5	27.36
27	157.6	82.1	78.2	44.93	41.92	24.37
28	235.3	216	73.6	84.93	4.06	79.13
29	172	100.7	159.3	52.86	43.39	-14.12
30	80.8	146.1	179.3	52.53	-28.82	-29.07

The previous presentation of embedded "object & metadata" as they result from the standardized use of test targets have been outlined sufficiently to develop an understanding of the relevance of these significant image properties. In view of the complex issues of digital long-term archiving, these facts can no longer be ignored. It would simply be irresponsible to continue to digitize image collections without documentation and technical verification options and not using this data. In the future, it would be like flying blind. If future society is to require trusted repository which will facilitate discussions about "fake" images and alike, it will be essential to produce, describe, and store trustworthy digital datasets, especially if focusing on emerging technology such as AI. With regard to cultural heritage, "nice looking images" are no longer sufficient.

This raises the question of how this quality assurance measurement data should flow (automatically) into the ingest process. Such approaches are conceivable with Metadata Encoding and Transmission Standard aka. "METS-containers"; but this is a different technical discussion. The basic prerequisite for this is a commitment to standardized quality assurance with one of the standards described.

6. Balance Sheet and Prospects

The mandatory use of standards such as Metamorfoze opens new perspectives for digital long-term archiving. These are not only limited to the ingest process, but migration steps can also be checked for their success using test targets. Using the printed device-independent LAB colour values, which document the specific properties of the respective colour patches, it will be possible to create future systems that can analyse and compare these values, regardless of future computer systems.

The fear that the digitization process and the long-term archiving of data will not become "simpler" due to the methods outlined is understandable. If the complexity of digital long-term archiving is to be reduced, this requires a paradigm shift that affects the entire process chain. If archiving is not placed at the end of this chain, but before communication, many things can be simplified. This requirement is called "Archive First". Here, the data should be indexed after digitization and then archived in the long term as part of an OAIS-compliant solution. It is then much easier to transfer the data to the various portals. The Pestalozzianum Foundation's comprehensive project is an impressive proof of this. All digitization and indexing requirements were primarily based on the archive perspective. Once the images have been ingested, they can be harvested relatively easily via standard interfaces such as OAI-PMH and published in any presentation system which handles the OAI-PMH protocol.

In the long term, the significantly higher digitization costs associated with this increase in quality will pay off. It forces institutions to adopt a "do it once, but do it right" approach. This ensures quality and security, while enabling work with transparent and traceable documentation in the future, which is what makes the paradigm shift to "Archive First" possible in the first place.

References

Dormolen, Hans van: Metamorfoze Preservation Imaging Guidelines 1.0. Image Quality, version 1.0. metamorfoze R national programme for the preservation of paper heritage 2012. Available online at https://www.metamorfoze.nl/sites/default/files/documents/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf, last checked on 18.02.2024.

FADGI (2020): OpenDICE and AutoSFR. Available online at http://www.digitizationguidelines.gov/guidelines/digitize-OpenDice.html, last updated on 25.05.2020, last checked on 18.02.2024.

HARVEY, Phil (2020): ExifTool. Read, Write and Edit Meta Information. Available online at https://exiftool.org, last last checked on 11.05.2020.

International Image Interoperability Framework (2020): IIIF. Available online at https://iiif.io/, last updated on 21.05.2020, last checked on 18.02.2024.

International Press Telecommunications Council (IPTC) (2020): Photo Metadata. Available online at https://iptc.org/standards/photo-metadata/, last updated on 25.05.2020.000Z, last checked on 25.05.2020.

iQ-Analyzer (2020). Available online at https://www.image-engineering.de/products/software/376-iq-analyzer, last updated on 25.05.2020, last checked on 18.02.2024.

ISO/TS 19264-1:2017 (2020). Available online at https://www.iso.org/standard/64221.html, last updated on 25.05.2020, last checked on 18.02.2024.

JHOVE. Open source file format identification, validation & characterisation. Available online at http://jhove.openpreservation.org/, last checked on 25.05.2020.

Library of Congress (2020): Recommended Formats Statement. Preservation, last updated on 25.05.2020, last checked on 25.05.2020.

NEUROTH, Heike; LIEGMANN, Hans; OSSWALD, Achim; SCHEFFEL, Regine; JEHN Mathias; STRATHMANN, Stefan (2010): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Hg. v. Heike Neuroth, Hans Liegmann, Achim Oßwald, Regine Scheffel, Jehn Mathias und Stefan Strathmann. Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland (V2.3). Available online at http://nestor.sub.uni-goettingen.de/handbuch/index.php, last updated on 2008, last checked on 18.02.2024.

PFEIFFER, Michel (2015): Wie können Bildbestände bewertet werden? Auswahl-, Erhaltungs- und Vermittlungsstrategien im Rahmen von Digitalisierungsprojekten. In: Zeithistorische Forschungen/Studies in Contemporary History (H2), S. 317–325.

PFEIFFER, Michel (2020): Wie digitalisiert man Bilder mit Blick auf die digitale Langzeitarchivierung? In: Information – Wissenschaft & Praxis, Band 71: Heft 5-6, S. 303-312. Available online at https://doi.org/10.1515/iwp-2020-2115, last checked on 18.2.2024.

PFEIFFER, Michel; BOSCHE, Anne (2023) Zugänglichkeit und Nutzbarkeit eines digitalen Kulturerbes in Bibliotheken und Archiven. Das Beispiel der Kinderzeichnungen der Sammlungen Pestalozzianum, S. 162. In: Schriften der Informationswissenschaft 76, Nachhaltige Information – Information für Nachhaltigkeit, Proceedings des 17. Internationalen Symbosiums für Informationswissenschaften (ISI 2023). Available online at https://zenodo.org/records/10009338, last checked on 18.02.2024.

Picturae Nederland (2020): Delt.ae. Online image quality assessment. Available online at https://deltae.picturae.com/index/home, last updated on 25.05.2020, last checked on 30.12.2023.

RIEGER, Thomas et al. (2023): FADGI Federal Agencies Digital Guidelines Initiative. Technical Guidelines for Digitising Cultural Heritage Materials. Creation of Raster Image Files. Available online at https://www.digitizationguidelines.gov/guidelines/FADGI%20Technical%20Guidelines%20for%20Digitizing%20Cultural%20Heritage%20Materials_3rd%20Edition_05092023.pdf, last checked on 18.2.2024.

Standard of Japan Electronics and Information Technology Industries Association (2002): Exchangeable image file format for digital still cameras: Exif Version 2.2. JEITA CP-3451. Available online at https://www.exif.org/Exif2-2.PDF, last checked on 25 May 2020.

Image Science Associates (2020). With the collaboration of Don Williams. Available online at http://www.imagescienceassociates.com/, last updated on 25.05.2020, last checked on 18.02.2024.