

Metadaten und Bilder

Christoph Rinne*

16. Dezember 2021

Inhaltsverzeichnis

1	Allgemeine Information	1
1.1	Vorbemerkung	1
1.2	Bildformate	1
2	Metadaten	2
2.1	IPTC	2
2.2	EXIF	3
2.3	DAI XML-Datenstruktur?	3
3	Software	3
3.1	XNView	3
3.2	exiv2	3

1 Allgemeine Information

Im Kern geht es um Metadaten und deren Bedeutung für die Nachnutzung von digitaler Objekte. Fotos, oder allgemeiner Rasterbilder, sind hier nur das gewählte Medium.

1.1 Vorbemerkung

Bilder - und es geht hier um Rasterbilder bzw. Fotos - haben durch die digitale Fotografie einen enormen Umfang erreicht, diese stellen in der Anzahl und vor allem im Datenvolumen bei fast allen (Forschungs-) Projekten den größten Bestand dar. Zwar gibt es eine immer bessere Bildanalyse und mit trainierten Neuronalen Netzen werden heute erstaunliche Ergebnisse erzielt (z.B. [YOLO](#)). Dennoch macht es Sinn, die bekannten Informationen zu den Bildern nicht (nur) in getrennten Listen zu führen, sondern in die Bilder hineinzuschreiben. Mit Blick auf die Nachnutzung zu dem vorgenannten noch das Problem der unzähligen Bildformate: von den RAW-Dateien der jeweiligen Kamera über JPG, TIFF und PNG als den bekanntesten zu weiteren Software spezifischen Formaten und Varianten innerhalb der Formate.

Automatisch eingebundene Metadaten in allen digitalen Dokumenten können viele Informationen zum Urheber der Datei transportieren. Sie sollten diese Daten also bewusst einsetzen und kontrollieren.

Leider kann über Metadaten auch Schadcode transferiert werden ([stackoverflow](#)).

1.2 Bildformate

Rasterbilder sind in ihrer Grundstruktur einfach, sie entsprechen einer Matrix in der jeder Zelle ein Wert zugewiesen wird, z.B. von 0 bis 255 (Schwarz bis Weiß). Ein Bild kann mehrere Ebenen besitzen (RGB) oder

*Christian-Albrechts Universität zu Kiel, crinne@ufg.uni-kiel.de

auch eine Farbtiefe von mehr als 256 (28) Farbwerten besitzen (224, true color). Derart gestaltete Bilder werden schnell sehr groß, weshalb diverse Kompressionsalgorithmen entwickelt wurden (JPG oder TIFF mit LZW). Im Fall von JPG werden dabei die Bilder unwiederbringlich verändert und qualitativ schlechter. Eine andere Strategie besteht in der Indexierung der wirklich verwendeten Farben (Farbpaletten-Modus, PNG). Eine gute Einführung und Überblick in das Thema Formate und Archivierung sowie weiterführende Links bietet [IANUS-FDZ](#). Als geeignet für die Archivierung gelten unkomprimierte TIF, GeoTIF, DNG. Wegen der sehr eingeschränkten Metadaten ([W3Org](#)) gilt PNG nur bedingt als geeignet und JPG gilt wegen der Kompression als ungeeignet. Wegen der hohen Verbreitung und der deutlich geringeren Größe ist JPG aber *de facto* ein Standard geworden. Um die Qualität von JPG nicht unnötig zu verschlechtern sollten sie wiederholtes Speichern vermeiden und bevorzugt mit verlustfreien Funktionen, z.B. der Rotation um 90° arbeiten. Auch das Schreiben von Metadaten in JPG erfolgt verlustfrei.

2 Metadaten

Bei Fotos gibt es zwei wichtige und schon lange existierende Gruppen von Metadaten mit jeweils definierten Feldern: IPTC und EXIF. IPTC zielt eher auf Urheberrechte und Bildinhalt, während EXIF mehr technische Informationen umfasst. Einige Felder, u.a. Datumsangaben, sind dabei redundant vorhanden. Bei XMP handelt es sich um eine allgemeine auf XML Standard basierte Definition von Metadaten, die dann mit Werten gefüllt werden können. XMP wird allgemein verwendet, u.a. für PDF und Bilderformate.

2.1 IPTC

IPTC steht für *International Press Telecommunications Council* und den Datenstandard, der zusammen mit der *Newspaper Association of America (NAA)* als *Information Interchange Model (IIM)* entwickelt wurde. Einen guten Überblick und Einstieg bietet die [Erläuterung zu dem Programm exivtool](#). Trotz der Definition kommen unsauber geschriebene Daten vor (u.a. Länge der Felder, Speicherort des IPTC-Blocks). IPTC Daten werden in Blöcken (*envelope records*) organisiert und sind innerhalb dieser erneut nummeriert (*tag id*). Beachten Sie bitte, es gibt ein IPTC Core Schema und ein IPTC Extension Schema. Eine kleine Auswahl der oft genutzten Felder aus den *IPTC Application Record Tags* sehen sie hier:

tag id	tag name	writable
80	By-line	string[0,32]+
85	By-lineTitle	string[0,32]+
90	City	string[0,32]
92	Sub-location	string[0,32]
95	Province-State	string[0,32]
100	Country-PrimaryLocationCode	string[3]
101	Country-PrimaryLocationName	string[0,64]
103	OriginalTransmissionReference	string[0,32]
105	Headline	string[0,256]
110	Credit	string[0,32]
115	Source	string[0,32]
116	CopyrightNotice	string[0,128]
118	Contact	string[0,128]+

Nicht alle Tags werden von jeder Software verwendet oder deren Bezeichnung sogar angemessen oder identisch übersetzt. Eine kleine Auswahl aus älteren Programmversionen sehen Sie hier:

IrfanView 3.97	exiv2 0.18.2	photoshop elements 3.0	XnView 1.96.2
Copyright	Copyright	CopyrightInformationen	CopyrightVermerk
Caption	Caption	Beschreibung	Objektbeschreibung
Caption writer	Writer	Verfasser der Beschreibung	Verfasser

IrfanView 3.97	exiv2 0.18.2	photoshop elements 3.0	XnView 1.96.2
Headline	Headline		Überschrift
Special instructions	SpecialInstructions		Sonstige Hinweise
Keywords	Keywords	Stichwörter	Sichwörter
Category	Category		Kategorie
Supplemental categories	SuppCategory		Zusätzliche Kategorien
Urgency	Urgency		Dringlichkeit
Byline	Byline	Autor	Fotograf/Ersteller
Byline title	BylineTitle		Titel des Fotografen/Ersteller
Credits	Credit		Bildrechte
Source	Source		Quelle
Object name	ObjectName	Dokumenttitel	Objektname
Date created	DateCreated		Erstellt am (JJJJMMTT)
City	City		Stadt
ProvinceState	ProvinceState		Bundesland
Country	CountryName		Land/Staat

2.2 EXIF

2.3 DAI XML-Datenstruktur?

3 Software

3.1 XNView

3.2 exiv2