

Metadata: Where and what to save

This article is ment for amateur/hobby photographers in first place and should help to find the way thru metadata jungle.

Character sets

I believe, you should make you familiar with basics of character sets (charset) before you decide to populate metadata.

ASCII

Complete ASCII charset contains 256 (0..255) characters. First 32 (0..31) characters are so called "control" characters, which can't be directly accessed by user. Next 95 characters (32..126) represent latin alpha-numeric characters and few punctuation characters:

!	"	#	\$	%	&	'	()	*	+	,	-	.	/	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>		
?	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^
_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~

Keep in mind, that above characters are the same on every PC, no matter what language (charset) is used.

The remaining characters (127..255) depend on PC's regional settings. So, for example, appereance of character 227 will depend on charset being used, for example:

- United States: π
- Western Europe: Ó
- Central Europe: Ě

Meaning, if you use some character not shown in table above, then you can't be sure that it will be shown correctly on any PC.

ANSI, Windows, etc.

To put it simple: these charsets are "extended" versions of ASCII. For user, the only important difference is, they contain more (regional and other) characters.

It is very important to know, that ANSI characters 32..126 are the same as in ASCII. However, portability limitation is almost the same as with ASCII: when non-ASCII characters (also called "foreign" characters) are used, then there's no guarantee, that these characters will be shown correctly on every PC.

UTF-8

UTF-8 charset (sometimes written as UTF8) is charset transformed from Unicode charset. There exist many transformations from Unicode, but in metadata case, UTF-8 is to be preferred, because (cited from Wikipedia): **UTF-8 is an 8-bit variable-width encoding which maximizes compatibility with ASCII.**

What does that mean in practice? It means, that if UTF-8 charset is used:

- non-ASCII (foreign) characters will be shown correctly on every PC,
- characters 32..126 (see table above) are written automatically as ASCII characters (again: shown correctly on every PC).

-these two properties should have big impact on our decission about where to save our metadata.

Metadata sections inside image file

Exif section

When photo is taken, camera automatically writes many metadata into image file. Most of that data, known as "Exif metadata", describes technical aspect of image: what camera/lens was used, what settings were applied, etc. Most recent cameras allow to store camera owner's name or/and copyright notice, which is then automatically written into each image file -and in most cases, that's it.

Exif data has great value for studying photography, but is less interesting for archiving purposes. With one (or two) exceptions:

- Inside Exif, **Date and Time** of when photo was taken, is automatically saved by camera.
- Inside Exif, **GPS** data of where photo was taken, is (should be) saved.

Now, there do exist some "interesting" tags in Exif:

- **Artist**, **Copyright**, **Software**, **OwnerName**, **UserComment**, etc.

-which are ment to be populated by user.

However, before adopting them, we must be aware of Exif limitations:

1st limitation: Officially, Exif supports ASCII characters only. However, **Metadata Working Group** (MWG) organisation

recommends using UTF-8 charset in Exif. No problem with that, as long characters 32..126 are used. But as soon we use some foreign characters, then again, there's no guarantee, that these characters will be displayed correctly if another software is used -software might "expect" that characters conforms Exif specification (are ASCII only) and will thus decode foreign characters wrong!

Something similar is already happening: To make use of foreign characters possible, today, most software uses ANSI charset when writting/reading Exif section. So, as long software makers don't change that, your foreign characters (encoded in UTF-8) won't be shown properly.

Ok, there's one exception, where this limitation doesn't apply: [UserComment](#) tag. For this tag, foreign characters can be used officially, but I wouldn't say this makes a big difference in Exif usage.

Conclusion: if you're from Poland, you shouldn't put your name in Exif, because your friend in Germany might not be able to read it.

2nd limitation: Exif doesn't have that many "interesting" tags as one might think. For example, in Exif, you **can't** save:

- location names of where photo was taken (city, country,..)
- names of peoples on photo
- keywords about photo content
- ...and many more.

Conclusion: if you wish to save more complete data about photo, then you're forced to look elsewhere.

IPTC section

In first place, IPTC metadata section was made for archiving purpose. It specifies many "about" (photographer, photo content, etc.) tags, which are ment to be populated by user.

At first, IPTC also allowed using ASCII/ANSI characters only, but now, Unicode/UTF-8 can be officially used as well. Of course, IPTC section has limitations:

1st limitation: Officially, tags defined in IPTC section are length limited. Some tags can only contain 3 characters (i.e. [Iptc:Category](#)), while other can contain several hundreds characters (most tags are limited to 32 characters, though). That's officially. However, in most cases, more than "allowed" characters can be (and many times are) written into IPTC section, and most software will show them all. But the fact remains: officially, limitation exist.

2nd limitation: Being a bit "old" standard, IPTC section doesn't specify tags we wish to have and need today. For example, there's no place, where you could save "rating" of your photo. The same is true for (photographed) people names, etc.

3rd limitation: [IPTC metadata specification for IPTC section isn't maintained anymore -instead, IPTC organisation decided to move IPTC metadata specification into XMP section.](#)

This fact added some confusion among photographers... Anyway, with many software today, when entering "IPTC" metadata, data actually isn't written into IPTC section only (or not at all) -by most "up to date" software, it is (additionally) saved into XMP section.

To sumarize: Above limitations are only valid for metadata inside IPTC section. That is, in case you use ExifTool command like:

```
exiftool -Iptc:City=Paris -Iptc:By-line="My name" ...
```

..values will be written into "old" IPTC section -because metadata section is specified. Ok, [By-line](#) tag only exist in Iptc, but you get the idea.

Conclusion: Old IPTC is dead... time to move on.

XMP section

XMP metadata specification defines how metadata is organized inside XMP section. XMP section can contain "any" metadata as long it follows the rule. There exist many specifications, but most known inside XMP section are:

- Adobe metadata (Photoshop imaging, pdf documents, etc)
- and
- IPTC metadata (for photography in general)

Now, believe it or not, XMP has no limitations. Limitless... sounds good, huh? Well, not neccessary. The problem I see is, in near future, XMP content can/will become huge and messy. Well, that's the price for having flexibility...

Note: In this article it is assumed, that XMP metadata is stored inside image file. I say that, because XMP can also be saved as separate (sidecar) file. In both cases however, data structure is the same.

There's another thing that might lead to confusion: it's quite hard to differentiate data between various metadata groups inside [XMP section](#). Let me (hopefully) explain:

Even **City** is "general" tag, it is saved into **XMP-photoshop** group. It's, I assume, because Adobe was the first defining that tag -I have no problems with that. But, on the other hand, **Rating** value (again, written by Adobe products) is saved into **XMP-xmp** group.

Now... By using **XMP-xmp** group, am I writing **Rating** into group I can "trust" for future? I mean, **XMP-xmp** group contain many tags; among others, there's **Author** tag, which is already marked as "non-standard".

Now, if that **Author** tag shouldn't be used, which tag to use then? How about **Author** tag specified in **XMP-acdsee** group? How to be sure my tags will be recognized properly? This question has it's place, because I believe, that my personal (about photo) data doesn't belong into "software specific" group. Does it matter, you ask? Remember **iView MediaPro** software? There, **XMP-mediapro** group was used to store **Event, Location, People,...** values. Today, all those tags are actually deprecated -similar tags in other groups are recommended to be used instead.

Let us return to **City** tag, which is (as said), defined inside **XMP-photoshop** group. Is that the tag to be used for saving city name? Not necessary... Not at all, if you wish to be more specific (and up to date with metadata). The thing is, it's not clear what that **City** tag (inside **XMP-photoshop** group) means: is that the city shown on photo, or is that the city from where photographer took the photo?

Conclusion: There's no doubt, XMP section is the place where you should store your data. The only thing you should do when starting is, choose the "right" group of metadata tags -you don't want to move your data from one place into another every two years, do you?

It depends on how much metadata you need/wish to manage, but right now, I would recommend using following groups:

- **XMP-dc** -for your name, copyright notice, photo title, keywords, etc.
- **XMP-iptcCore** -for your contact data (address, mail, phone, etc.); in short, stuff ment for pro photographers.
- **XMP-iptcExt** -for location data, event notice, names of persons on photo, etc.

Of course, there might be specific need to also use some other group. In most cases however, above groups will suffice.

We have mentioned, that inside XMP tags may exist, which have equal name: **Event** tag, for example. If you use ExifTool for writing:

```
exiftool -Xmp:Event=Birthday MyPhoto.jpg
```

..then value will be written inside both: **iptcExt** and **mediapro** group -because both have this tag specified. If you wish to be precise about where exactly to write, then specify the group:

```
exiftool -Xmp-iptcExt:Event=Birthday MyPhoto.jpg
```

..which is the way I recommend.

How many metadata can you handle?

Writing metadata into all sections: Exif, IPTC and XMP?

Just to be sure, that data will guaranteed be seen by any software, some are saving the same metadata values wherever possible. Do what you think is the best for you, but speaking for me, it's pure waste of time. No matter how perfectly you think your "workflow" is, at the end (as your photo collection grows), you will give up doing this.

DateTime of when photo was taken is typical case... As mentioned above, this data is automatically saved by camera into **Exif** section. Now, I really see no reason to have the same value elsewhere too (inside **iptc** or/and **Xmp**, for example).

Ok, let's say, you have the reason for doing this. But do you know how photo software handles this? In 99% cases it goes like this:

1. **Look into Exif:** if **DateTime** exist here, then show that value and don't look elsewhere. If not, go to 2.
 2. **Look into iptc:** if **DateTime** exist here, then show that value and don't look elsewhere. If not, go to 3.
 3. **Look into Xmp:** if **DateTime** is defined here, then show that value.
- you get the point: if value is defined in **Exif**, then there isn't of much use to have the same value elsewhere.

Next very typical case is **Exif:Artist** tag, which is equivalent to **iptc:by-line** and **Xmp-dc:Creator** and follows the same rule as described for **DateTime** above.

And then, there are keywords, which can, again, be stored inside **iptc** or/and **Xmp** -here, I really have no idea how to keep them synchronized... etc.

And finally, for photos taken with digital camera, here's my personal view on where and what to save.

What goes into Exif section

Artist and Copyright

It's a good idea to save this data inside Exif. Because, in future, it might happen you will reorganize/rewrite (or even delete) complete **Iptc/Xmp** section -having both inside Exif, you won't need to worry about them. Btw. some cameras can automatically write that data in Exif.

However, I don't say that data "must" be in Exif (only). It's a personal decission and if you decide to manage all your

"personal" metadata in one section only (i.e. in Xmp section), that's fine too.

[DateTimeOriginal](#) and [CreateDate](#)

If photo was taken with digital camera, then both values are already inside Exif. Eventually, you will modify these two values in case date/time on camera was wrong at the time when photo was taken. And why are two DateTime tags there (having equal value)? Because, for scanned photos, their values can't be equal.

And what about [ModifyDate](#)? Don't ask me why, but some are desperate trying to keep all three DateTime values equal. I mean, what does "modify date" mean to you? Exactly!

[GPS](#) data

If at all, then [GPS](#) data should always be written into Exif. In this case, at least [GPSLatitude](#), [GPSLatitudeRef](#), [GPSLongitude](#) and [GPSLongitudeRef](#) values should be written -when doing this, ExifTool will automatically set obligatory [GPSVersionID](#) value.

And that's all what photographer "need" to write into Exif.

[What goes into Iptc section](#)

Nothing. I simply see no reason to start using this section anymore. Reading many forums, I can see, that many are magically attracted when hearing "iptc". But reading further, it many times reveals, that many don't really distinguish between Iptc metadata [section](#) and Iptc metadata [standard](#). So, I repeat here again:

[IPTC \(organization\) decided to move specification into Xmp section](#) -by making this move, Iptc [section](#) became obsolete.

[What goes into Xmp section](#)

There's nothing we can't save here -but saving everything would be too much to ask. Thus, let's see what tags amateur/hobby photographer would eventually like to populate.

[Xmp-dc:Creator](#)

-here, photographer's name should be saved. This is Xmp equivalent of [Exif:Artist](#) tag.

[Xmp-dc:Rights](#)

-here, copyright notice can be saved. This is Xmp equivalent of [Exif:Copyright](#) tag.

[Xmp-dc:Date](#)

-here, date and (optionally) time can be saved. This is equivalent of Exif DateTime value(s). It is important to note, that here, "partial" DateTime value can be saved, i.e. "1978:06" -usefull in cases when exact time isn't known (i.e. old scanned photos). I mention that, because in Exif, complete (date and time) value is required.

[Xmp-dc:Title](#)

If you decide to title your photos, then it should be written here. Even there's no length limit, it's expected to be short. That is, no matter what you might think, but others are usually not that much interested to read long stories.

[Xmp-dc:Subject](#)

This is the place, where [keywords](#) should be saved. Tag [Subject](#) is "multi-value" tag, which means, it can hold multiple (internally separated) values. This tag can hold "normal" keywords, which means, all keywords have equal "weight"; i.e. if you store keyword "animal" and keyword "spider" into [Subject](#) tag, then both will have equal importance.

There also exist tags, which can hold keywords "hierarchically" structured. In this case, for example, keyword "animal" can be one of many main/root keywords, holding other sub-keywords, for example "spider", "bird", etc. If you're interested doing it that way, then take a look into [XMP section](#) and decide what tag you wish to use. But be warned: maintaining such keyword structure isn't necessary simple.

Keywording can be very helpfull later -but only if done right! Thus, before starting, make sure you know what "keyword" actually means and what benefits you expect by having them. Keywords is not something one must have in metadata. However, if you decide writting them, then in my opinion, it only make sense, if all your photos are "keyworded". I say that, because all this might take a lot of time.

[Xmp-dc:Type](#)

This tag has the same characteristics as tag [Subject](#) (above): it can contain multiple values. While [Subject](#) tag usually defines the content shown on photo (i.e. "bird", "sunset", etc.), tag [Type](#) is ment to tell about "kind" of photo. This can be, i.e. "portrait", "landscape", "studio", "sport", etc. But being multi-value capable, you can save "portrait" and "studio", or "portrait" and "outdoor" here, of course. That way, it might be easier to find i.e. all "portrait" photos taken in "studio". Similar to keywording: if at all, then start with only few (major) types -you can add more later.

[Xmp locations tags](#)

In most cases, we want so save location of photo content. That is, for example, if photo of village was taken from the hill nearby, then we want to save location of the village (on photo); not the location of the hill. In this case, the following

Iptc4xmpExt tags should be used:

Iptc4xmpExt

[Xmp:LocationShownCountryName](#)
[Xmp:LocationShownProvinceState](#)
[Xmp:LocationShownCity](#)
[Xmp:LocationShownSublocation](#)

Adobe legacy

[Xmp-photoshop:Country](#)
[Xmp-photoshop:State](#)
[Xmp-photoshop:City](#)
[Xmp-IptcCore:Location](#)

I have already mentioned multiple tags inside Xmp section, which serve the same purpose and locations tags are no exception.

Right now, most software will write locations to (and read from) tags introduced by **Adobe** and it's hard to predict if this "habbit" will change soon. I mean, Adobe was here first and **IPTC** was simply too slow with Xmp **IPTC Extension** implementation. Inbetween, zillions of photos are already tagged...

On the other hand, Iptc tags in [Xmp-iptcExt](#) group are "standardized" and recommended to be used. If you take a look there, you can see there are many other tags defined which might be usefull, i.e. the location from where photo was taken.

Usually, not all locations tags can be populated on all kind of photos. For example, if you make a trip across desert in Tunisia, you will hardly know later in what province exactly you were.

[Xmp-iptcExt:PersonInImage](#)

Here you can save names of persons shown on photo. Similar as [Subject](#) and [Type](#) tag, this tag can hold multiple values (names).

Maybe I should mention this: because only one tag having name [PersonInImage](#) is defined for Xmp, we don't need to specify Xmp group. That is, in this case, both commands

```
exiftool -Xmp-iptcExt:PersonInImage=Frankie myPhoto.jpg
```

and

```
exiftool -Xmp:PersonInImage+=Johnny myPhoto.jpg
```

-will populate the same tag. The same goes for every tag, which has unique name inside Xmp.

[Xmp-iptcExt:Event](#)

As you might guess, here you can write "birthday", "fishing", etc. Here, however, you shouldn't use "short" group name (i.e. [-Xmp:Event="race"](#)), because this tag is defined elsewhere!

That is, exact group must be specified in this case, i.e. [-Xmp-iptcExt:Event="race"](#).

Summary

I think, now is the time to rethink few things, before you jump populating metadata. As you can see, there are tags, that serve quite similar purpose. For example: you can decide to write the words like "birthday" into [Subject](#) tag, or into [Event](#) tag. And no, nobody says you must use both tags. That is, use Event tag, if you think, your photography interests can be classified by "events" -otherwise, use [Subject](#) tag only.

What I'm trying to say is, you dont need to populate all tags I've mentioned above! And before everything, first start with few "must have" tags; i.e. locations tags. When you finish this on all photos you have, you will get a feeling how much time all that requires.

Bogdan Hrastnik
 February, 2012