About Photo Metadata

Considerations and advice shared by
Michael Steidl
Lead of IPTC Photo Metadata Working Group
at the IPTC Photo Metadata Conference
13 October 2020

Who is the PMD Working Group

A group of photo metadata people from an IPTC member (in random order)

David Riecks, PLUS
Karl Csoknyay, Keystone-SDA
Brendan Quinn, IPTC MD
Michael Steidl (Lead)
Annette Feldman, AP
Linda Burman
Jeff Sedlik, PLUS
Alfonso Gutierrez, CEPIC
Margaret Warren
Scott Eells, Bloomberg
Alison Sullivan

© 2020 IPTC (www.iptc.org) All rights reserved
Let’s talk about …

Photo Metadata (PMD)

MUST
should
can
... may
does not
tell the truth

Clarification

Photo Metadata telling the truth

= True and correct facts about the image

(... and we don’t talk about the truth of the pixels)
The Metadata Sea and its Photo Metadata Bay

Let’s investigate it

The Islands in the Photo Metadata Bay

Homebase
Photographer & Company

Publishers

Agencies & Suppliers

Viewers
The Goal of Photo Metadata: Communication

PMD communication reality check
PMD’s fate in the communication

Homebase island

= the creator(s) of the photo, the photographer and her/his company

Metadata is based on knowledge very close to the pictured event or object, it is checked and written down.

Other involved parties:

- Camera makers
- Makers of imaging software supporting photo metadata standards (e.g. Lightroom, Photo Mechanic, Fotostation, Photoshop …)
- Makers of Digital Asset Management software supporting photo metadata standards
- Providers of Artificial Intelligence/Machine Learning systems supporting images
Agencies/Suppliers island

= companies collecting, aggregating, supplying and licensing photos

Metadata may be edited: let’s assume corrected or metadata required for attracting customers (e.g. keywords) added.

Other involved parties:

• Makers of imaging software supporting photo metadata standards
• Makers of DAM software supporting photo metadata standards
• AI/ML system providers supporting images

Publishers island

= organisation posting licensed photo(s) on the web (or in print)

Metadata is used, replaced or removed as the organisation is interested in.

Other involved parties:

• Makers of editorial systems for journalists
• Web Content Management system makers
• Content Delivery Networks reformatting photos
• Search engines crawling the website and showing found images in search results … with metadata
Viewers island

= persons viewing a photo and reading metadata information (if readable)
Metadata must be displayed to them – else they can’t get the metadata information.

Other involved parties:
• Makers of software showing the photo:
  Image viewers, web browsers, PDF viewers, file metadata viewers
• Search engines showing photos as results – let’s hope - with metadata

The Goal of Photo Metadata: Communication

... and many are involved
How to communicate photo metadata well

• This communication needs a language – beyond single words
• A language needs a grammar, rules for building sentences making a statement:

Most languages of Europe and the Americas use this basic grammar …

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Frank”</td>
<td>“takes”</td>
<td>“a photo”</td>
</tr>
<tr>
<td>“Francois”</td>
<td>“prend”</td>
<td>“une photo”</td>
</tr>
<tr>
<td>“Franz”</td>
<td>“schießt”</td>
<td>“ein Foto”</td>
</tr>
<tr>
<td>“Francesco”</td>
<td>“toma”</td>
<td>“una foto”</td>
</tr>
<tr>
<td>“Franco”</td>
<td>“scatta”</td>
<td>“una foto”</td>
</tr>
<tr>
<td>“Frank”</td>
<td>“robi”</td>
<td>“zdjęcie”</td>
</tr>
</tbody>
</table>

Many standards for photo metadata follow this basic grammar and build statements this way:

• The Subject is “this image”
• The Predicate is a field/property/tag defined by a (photo) metadata standard
  Such a predicate defines what this statement tells about the photo!
• The Object is the value of a field set by a metadata editor

Example: Subject=“This image”, Predicate=“has Creator”, Object=“Carola Ranotti”

Photo metadata fields tell simple sentences, make statements for the human viewer.
Dear Photo Metadata Standard User …

… take care to use a metadata field as defined by its standard
This supports

For IPTC Photo Metadata this helps:
• **The User Guide**, grouping fields
by these categories

… and giving hints for their use.

• **The Standard document** with a definition of each field (= property)

E.g.:

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Contains the name of the photographer, but in cases where the photographer should not be identified the name of a company or organisation may be appropriate.</td>
</tr>
</tbody>
</table>

Photo Metadata Standards out there

… they provide a well defined vocabulary of fields = predicates of communication

<table>
<thead>
<tr>
<th>Exif</th>
<th>IPTC PMD Standard</th>
<th>schema.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date + Time Original (GPS data)</td>
<td>Date Created</td>
<td>(Use an ImageObject)</td>
</tr>
<tr>
<td>Artist</td>
<td>Location Created</td>
<td>dateCreated</td>
</tr>
<tr>
<td>Copyright</td>
<td>Creator</td>
<td>locationCreated</td>
</tr>
<tr>
<td>User Comment And:</td>
<td>Copyright Notice</td>
<td>creator</td>
</tr>
<tr>
<td>Height, Width Orientation Exposure time, …</td>
<td>Description</td>
<td>(copyrightNotice as draft)</td>
</tr>
<tr>
<td>And: Headline Location Shown Credit Line, …</td>
<td>And: caption</td>
<td>And: headline</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contentLocation, …</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(creditText as draft)</td>
</tr>
</tbody>
</table>
Use of standards in a photo workflow

<table>
<thead>
<tr>
<th>Metadata presets</th>
<th>Taking the photo</th>
<th>Edit metadata /1</th>
<th>Edit metadata /2…</th>
<th>Photo in web page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata default values are set in the camera by its user.</td>
<td>Metadata of the camera and of a preset on the memory card is embedded into the image file.</td>
<td>Exif Tag values are copied to IPTC fields. Metadata fields are edited, IPTC fields and some Exif Tags are updated &amp; embedded.</td>
<td>Metadata fields are edited, IPTC fields and some Exif Tags are updated synchronously and embedded.</td>
<td>Exif and IPTC metadata should stay embedded, its values can be copied to schema.org metadata properties.</td>
</tr>
<tr>
<td>Exif Tags can be set</td>
<td>Exif data is generated and embedded</td>
<td>Exif data is copied to IPTC; may be updated</td>
<td>Exif data is updated (in sync with IPTC)</td>
<td>Exif data should not be removed</td>
</tr>
<tr>
<td>IPTC fields may be set</td>
<td>IPTC presets can be copied and embedded</td>
<td>IPTC fields may get copy of Exif values; is updated</td>
<td>IPTC data is updated (in sync with Exif)</td>
<td>IPTC data should not be removed</td>
</tr>
<tr>
<td>schema.org</td>
<td></td>
<td></td>
<td></td>
<td>If schema.org is used: IPTC and Exif values should be copied to Image Object properties</td>
</tr>
</tbody>
</table>

The threat: broken communication

Homebase
Photographers & Company

Agencies & Suppliers

Publishers

Why? Correct information is not available

Viewers
**Why? → Check is the input correct?**

- Photographer when taking the photo: is date and time of the camera correct, can GPS data be received properly?
- Photographer as first PMD editor: are all known data written down correctly?
- Photo metadata editor on a desk: are all values written by the photographer correct including spelling? (4-eyes principle)
  If a geo-location service is used with the GPS data: do the generated location names match information of the photographer?
- Keyword/term added by Artificial Intelligence (AI) or Machine Learning (ML): does the provided term match the content closely? (The "score" helps.)

**Why? → Check are the values kept alive?**

Keeping alive = removed only with good reasons and not removed inadvertently

- Check what metadata fields and values are embedded into an image file
  - Imaging software with photo metadata support can be used – but some pieces of software don’t support all IPTC fields.
  - IPTC suggests to use its Get IPTC Photo Metadata: [https://getpmd.iptc.org/](https://getpmd.iptc.org/)
- Check if metadata fields displayed and used by Google Images are embedded
- Check if embedded metadata of a photo posted on the web exists as expected:
  - Check your own website, in your web shop
  - Check the website of your customers
  (See what the IPTC Photo Metadata Explorer found on widely used news sites)
Why? ➔ Check are the values kept alive? … continued

- Check how open systems you are using for displaying photos treat photo metadata:
  IPTC has checked the fate of embedded metadata on 500px, Behance, EyeEm, Flickr, fotocommunity.de, Google Photos, Instagram and more in 2019, and published the results and how the sites were tested.

- Check your license contracts: do they tell customers to keep metadata alive? (… at least fields relevant for rights ….)
Helpers for getting out (1)

- The Photo Metadata Bay needs a guide for synchronizing metadata values across standards; also called mapping of fields.

Currently IPTC discusses this goal with CIPA, the authority behind Exif, and Schema.org

<table>
<thead>
<tr>
<th>Property Name</th>
<th>Definition</th>
<th>Authority: IPTC</th>
<th>Authority: Exif</th>
<th>Authority: Schema.org</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>Contains the name of the photographer. An empty value shall not be assigned to the name of a company or organization.</td>
<td>Person who created the image</td>
<td><a href="https://schema.org/creator">https://schema.org/creator</a></td>
<td><a href="https://schema.org/author/to-FrequentBy/IPTC">https://schema.org/author/to-FrequentBy/IPTC</a> (Roles)</td>
</tr>
<tr>
<td>Creator’s Contact Info</td>
<td>The creator’s contact information includes a telephone and address.</td>
<td>Person who created the image</td>
<td><a href="https://schema.org/creator">https://schema.org/creator</a></td>
<td>(Person properties)</td>
</tr>
<tr>
<td>Creator’s Job Title</td>
<td>Contains the job title of the photographer. It’s the role of a person or an organization rather than the role of the creator of the image. This is a free text field.</td>
<td><a href="https://schema.org/creator">https://schema.org/creator</a></td>
<td><a href="https://schema.org/jobTitle">https://schema.org/jobTitle</a></td>
<td></td>
</tr>
<tr>
<td>Credit Line</td>
<td>The credit line (and/or organizational entity) required by the subject of the image is to be filled by a metadata capable of using Creator’s Job Title.</td>
<td><a href="https://schema.org/creditLine">https://schema.org/creditLine</a></td>
<td><a href="https://schema.org/creditLine">https://schema.org/creditLine</a></td>
<td></td>
</tr>
<tr>
<td>Date Created</td>
<td>Date the data or the value of the content of the image was created/modified. Date of creation of the digital representation.</td>
<td>Date and time of original data generation</td>
<td><a href="https://schema.org/Temporarily">https://schema.org/Temporarily</a></td>
<td><a href="https://schema.org/created">https://schema.org/created</a></td>
</tr>
<tr>
<td>Description</td>
<td>A textual description including captions of the image.</td>
<td>Image title</td>
<td><a href="https://schema.org/title">https://schema.org/title</a></td>
<td><a href="https://schema.org/description">https://schema.org/description</a></td>
</tr>
</tbody>
</table>

Helpers for getting out (2)

- Makers of software supporting to edit photo metadata (for single computers, for a server-based system) should consider:
  - Keeping defined IPTC and Exif photo metadata values in sync (mostly done!!)
  - A function creating structured schema.org metadata for web pages.

- Makers of web content management systems (CMS) should consider:
  - Importing IPTC and Exif metadata correctly to their database
  - A function transforming embedded IPTC and Exif metadata to structured schema.org metadata for a web page showing the photo.
  - Not to remove embedded metadata without any notice
  - Not to remove relevant fields like Creator, Copyright Notice, Credit Line, Description, Web Statement of Rights (Copyright URL), Licensor URL

- Makers of tools (modules, libraries) for web sites should consider developing tools for an easy retrieval and display of embedded metadata.
Helpers for getting out (3)

• To improve the **delivery speed** of photos on the web is not a good reason for **stripping off all metadata** (“slim files are delivered faster”)
• Make metadata of photos on the web more transparent = display the fields
  ➤ Google Images makes them transparent in search results.
• CIPA is currently discussing with IPTC which **Exif Tags** set at the time of taking an image **shall be protected and conveyed endlessly**. Software makers will be encouraged to implement that.
• What about **making saved/embedded photo metadata secure** against not permitted changes?
  What about a **history of metadata changes** which cannot be destroyed?
  ➤ The Adobe Content Authority Initiative (CAI) has plans to achieve that.

Summary for photo businesses

➢ Take care to set correct values in the metadata fields
➢ Check if they are embedded correctly in your files
➢ Check if they are in photo files on any website controlled by you … and IPTC tools help you checking this.
➢ If not: select a better software
➢ … or talk to the maker of your software and request an improvement.
➢ Check if metadata gets lost somewhere in a supply chain you deliver to
➢ Remind customers not to remove metadata when publishing your photos
➢ Keep your eyes open for new things helping to keep metadata alive – and support this thing.
Let’s all be behind this goal

Photo Metadata tells the Truth

Homebase
Photographers & Company

Publishers

Agencies & Suppliers

Viewers