

IMAGE METADATA PLANNING

A guide for business users



Written for CEPIC by Sarah Saunders, [Electric Lane](#) with the help of [the IPTC Photometadata Working Group](#)
© CEPIC and IPTC All Rights Reserved

CONTENTS

1. Why metadata?
2. This guide
3. The business case for metadata
4. Entering and storing metadata
5. Ten metadata workflow principles
6. What software to use?
7. Workflow Issues
8. Keywording and retrieval
9. IPTC Core and Extension
10. New fields in IPTC Extension
11. Planning your metadata
12. Use this guide to help you plan
13. News agency sample workflow
14. Stock agency sample workflow
15. Museum sample workflow
16. Resources
17. Acknowledgements

Interactive IPTC field chart

IPTC Core fields and how they are used

IPTC Extension fields and how they are used

1. Why metadata?

We've always had metadata on our images, as caption, credit and copyright information on the back of a print or on a slide mount. Important copyright and identification data was firmly attached to the image.

In the online world there can be many copies of a single image, and with millions of images on the internet, metadata is essential for identification and copyright protection. We have to make sure this metadata travels with the image as a digital label, and remains with it over its lifetime.

The metadata associated with an image can provide information about

- Copyright and other rights associated with the image
- Contact information of copyright holders and licensors
- Image content
- Artworks, buildings and people portrayed in the image
- Search terms (keywords)
- Technical details of the photography
- Rights restrictions for use of the image
- Rights granted under a licence to use

Driving the image business

Metadata is the driver for the image business. Without it, images cannot be retrieved in a focussed way, their provenance is unknown, and they cannot be used legally.

Don't let your images become orphans

Images can become nameless orphans if there is no information inside the image file.

Data will always be needed

Although there are now visual methods of finding images, they cannot replace the detailed information available in metadata fields.



2. This guide

We hope the guide will provide information and guidance for businesses wanting to use metadata in their own day to day work.

Metadata drives images forward in an increasingly automated workflow. How it is used depends on internal workflow, the needs of customers, and software capabilities at both ends.

This guide sets out some of the principles which can be applied to planning metadata, with some examples of how this can be done.

In the CEPIC/IPTC metadata survey carried out in 2010, we found that 70% of image suppliers were using IPTC fields as specified, and that the fields used by the companies and their clients varied. More than 70% considered the use of IPTC fields to be very important to their business with a majority believing its importance would grow with time.

Some companies were unaware of the IPTC Extension fields, which were added to IPTC in 2008. We have discussed the new fields, and the gaps they fill in the IPTC schema, and hope this will help companies to use the full range of fields where these are relevant to their business.

[CEPIC](#) is the *Coordination of European Picture Agencies Stock, Press and Heritage*
[IPTC](#) is the *International Press Telecommunications Council*

Every company has its own workflow

In every sector we consulted there were varying ways of capturing and distributing metadata. The guide sets out to get companies thinking about their workflow and is not prescriptive except that we want to encourage use of the IPTC fields as specified.



3. The business case for metadata

Automation

Any business handling images needs to introduce automated systems where possible. Metadata schemas provide consistent ways of labelling images that drive automation, and allow images to be tracked, searched and used.

Beyond digitisation

Digitising images is just the start. Without effective metadata they cannot be found and used. In an image rich world, it is important to plan an image workflow including storing, archiving and retrieval. Metadata forms an integral part of this workflow.

Working within the law

Compliance – making sure that a company's activities operate within relevant laws and regulations – is becoming increasingly important. Companies recognise the risks and create administrative systems to protect themselves. Sound metadata workflow helps protect businesses from unpleasant surprises in the area of copyright, model rights and rights to use an image, and should form an integral part of compliance planning.

Capture data from the image file

Batch upload and processing of images is easier if existing metadata can be captured from the image file.

Plan data and digital workflow together

Ideally, planning for digitisation should go hand in hand with metadata and retrieval planning. However, it is never too late to start!

Respect IP and protect your brand

Creators and people portrayed in images are increasingly aware of their rights. It can be disruptive, expensive and bad for brand reputation if these rights are infringed or ignored.

3. The business case for metadata

Business Critical Metadata

Every business has to make its own assessment on how much metadata to add, and what value it brings. But critical metadata such as Image rights information and identification is *always* necessary, and should be part of any business image workflow.

Business advantage

Beyond this, good use of metadata increases competitiveness. Better access to images, faster retrieval and efficient rights handling are important success factors for businesses producing and distributing images, and also for those managing their own image collections.

Return on Investment

It costs money to acquire, handle, retrieve and distribute images using best practice guidelines and sound metadata. It costs a lot more to do the same job without effective metadata. If images are a valuable asset to a business, then it is worth investing in metadata.

Businesses which fail to invest in metadata will find they are at a disadvantage, both now and in the future. Most companies will want to run automated workflows, with effective asset retrieval and legal compliance built in. That's the way business is heading.

Plan ahead first

*Our interactive **IPTC field chart** shows examples of how metadata can be used. You can use the chart to plan your own workflow.*

..... then automate

"The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency."
Bill Gates

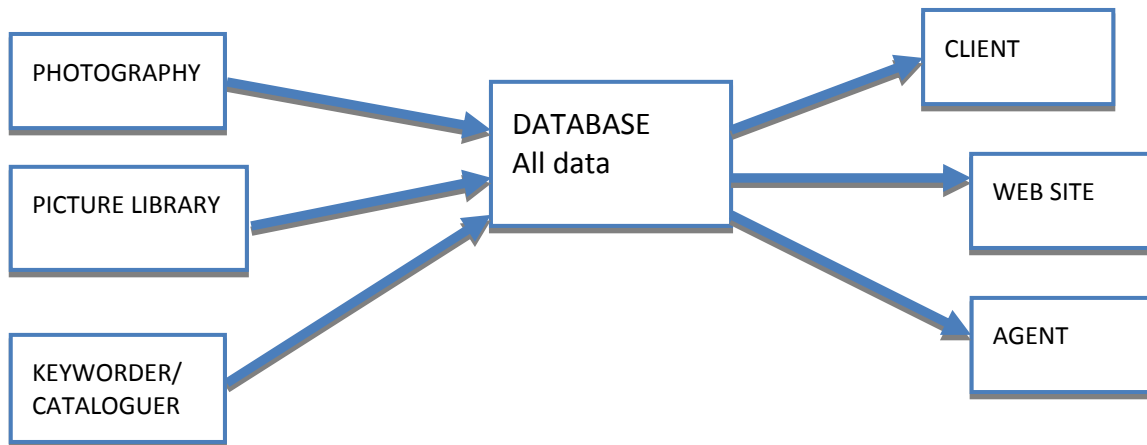
Whole organisation planning

It is useful to do a metadata audit across your entire organisation. Integrating metadata work across departments can reap substantial benefits.

4. Entering and storing metadata

The two main places to enter metadata are to embed it in the image file (as Exif, IPTC and PLUS) and to store it outside the image file in a database.

The IPTC fields carry workflow related data, while the database may also hold management data not required by people receiving or using the image. In general, the database is the repository for all the information a company holds about an image.



Exif is the Exchangeable image file format for files from digital cameras and other devices which produce digital images. The Exif format holds technical data recorded at the time the image was shot, such as shutter speed, aperture, date, resolution and image size.

IPTC (International Press Telecommunications Council) is a consortium of news agencies which has been setting standards for news exchange since the 1970's. The IPTC Photometadata schema is a standard for metadata held in the image file, and is widely used in the photography and news industries.

The PLUS Coalition is an international non-profit initiative whose mission is to simplify and facilitate the communication and management of image rights.

How is data entered?

By users

- Into the image file
- Into an image database

Automatic capture

- Data can be captured from camera or scanner

Data in

Data flows into the database

- from Excel sheets and CSV files-from other databases, photographers
- direct from XMP fields in the image file (IPTC, PLUS, Exif)
- direct data entry into the database at the user interface

Data out

Data flows out of the database

- to XMP fields in the image file (IPTC and PLUS)
- to information viewed on a web interface
- to Excel or CSV exports

CSV (comma-separated values) is a common format for data transfer. CSV files can be created and read in spreadsheet programmes like Microsoft Excel.

XMP is Adobe's Extensible Metadata Platform, an XML based labelling technology that allows data to be embedded in a file.

XML Extensible Markup Language is a set of rules for encoding documents in machine-readable form.

5. Ten Metadata workflow principles

1. Capture information once

Information should be entered once only, as soon as possible in the workflow.

2. Embed critical data in the image file

Information should be embedded in the IPTC and PLUS fields when images are distributed, including via the internet. Data carried separately from the image is likely to become detached. Copyright and identification information should *always* be retained in the image file.

3. Keep all data in your main image repository database

Your database will be backed up and should contain all data relating to the image, including management data that is only used internally.

4. Don't dump data - retain technical (Exif) metadata

Some technical metadata may seem irrelevant to your workflow now, but with increased use of automation it may become important later.

5. Plan your metadata workflow

Decide what data you need internally and for export to your customers, and decide who will enter the data (see our [interactive IPTC Field Chart](#) and sample workflows).

Creator stays the same

The photographer's name should be entered under 'Creator'. This will not change, even if copyright changes hands.

Low res files on the desktop

Picture researchers complain that low res images downloaded to their desktops, have no data with them. Help by embedding metadata in the image file.

Optimising images

The photo processing and printing industry is developing automated systems for image optimisation which use Exif data captured at image creation.

Keep legacy image numbers

The photographer's image number may be needed in the future. Keep it in your database.

Plan before you DAM

If you are buying a new image DAM system, plan your data workflow in advance. The best systems will handle data flexibly in and out of the database, importing and exporting XMP fields to and from the image file.

6. Check your legacy data

Watch for interoperability of datasets. Check for handling of diacritics, separators, date formats, and other inherited data features. It is best to have data checked by an expert.

7. Separate your data

Data should be stored in the most granular way possible. Data can be merged automatically, but separating data is difficult if there are no markers.

8. Use the correct IPTC field

Consult the [Interactive IPTC Field Chart](#) , the [IPTC Core and Extension User Guide](#).

9. Use automation where possible

Use bulk processing and scripts to handle large amounts of data. Scripts can manage data flow using logical rules.

10. Audit and check your data

You should set up quality control points and regular monitoring of metadata. Audits can be done automatically, using scripts, while manual spot checking of keywords, captions and other data is needed to ensure ongoing quality.

Data incompatibilities

Problems with data may be linked to incompatible character sets. Database software should be Unicode compatible, so it handles accents, umlauts and other character sets (e.g. Japanese, Chinese).

All in the caption?

Some image libraries put headline, caption, and copyright info in the description field so it can be picked up by news systems. Information should also retained in the relevant separate fields as well.

Link to IPTC

The guide for Adobe CS products is online at [IPTC-PLUS Photo Metadata Toolkit for Adobe CS](#).

Use IPTC panels for productivity

The new [IPTC-PLUS panels](#) for use in Bridge enable bulk import and export and copying between IPTC Extension and PLUS shared fields.

Secure your investment

If you invest in inaccurate or low quality data, it will affect your business in the future. Get it right now.

6. What software to use?

Photographers may enter data in

- Lightroom, Aperture, FotoStation and other image management software
- Photoshop, Bridge, Photo Mechanic and other browsers
- Spreadsheet programmes (Excel and others)

Picture agencies may use the above, but also

- Spreadsheet/CSV files (for import and export to data to agents)
- Picture library database

Publishers may also enter data in

- Enterprise DAM/management system

Organisations and companies may enter data in

- Image DAM systems
- Content Management
- Enterprise DAM systems

+ does your software ?

- Support XMP metadata
- Import, read, and export IPTC Core and Extension fields (reputable software should do this)
- Read IPTC-IIM legacy data
- Allow for flexible administrator managed import and export of IPTC data
- Allow you to do bulk data entry
- Allow for remote data entry by others (e.g. photographers)
- Read and use PLUS fields

- does your software ?

- Strip Exif, or IPTC-IIM or XMP metadata headers from incoming images
- Tie you to inflexible import and export templates
- Add invisible characters which may affect your data on export

7. Workflow Issues

How much metadata do we need?

That will depend on your own workflow and resources, and on what your customers need to know about the image. Start with a list of critical rights and identification data, and add descriptive and administrative data needed to run your business.

Who should enter the metadata?

Data needs to be added as early as possible in the supply chain. Every business needs to create its own data workflow (see [Planning Your Metadata](#)).

These people may add data

- Photographer
- News agency picture desk
- Picture Library staff
- Client

Some agencies ask photographers for contact data, captions, headlines and keywords. They may enhance that data in their own image database system. News agencies often issue templates to photographers with relevant shoot data, to keep photographers' work to a minimum in time critical situations.

How much data?

- *Entering data takes time, but some data can be captured automatically (e.g. GPS location data).*
- *More granular data will make your images easier to retrieve .*
- *Although agents, distributors and clients may not handle all fields now, they will in the future. Enter data in the correct field to future-proof your content.*

Remember

- *The picture library is the central metadata entry point. Make sure someone is assigned to enter, check and manage the data.*
- *Templates can aid in bulk data entry. The [IPTC panels for Bridge](#) hold templates for data entry and other workflow aids.*
- *Be realistic about who will enter data – it takes time.*
- *Get everyone on board to recognise the benefits of good data.*

8. Keywording and retrieval

Finding Images

There are two main forms of search:

- *Free text* - where all words in a text field are searched. For example, the words in a caption text are searched.
- *Keyword search* – where specific words are assigned to an image, specifically for search purposes.

Search systems can use a combination of these, and have different ways of weighting search results.

Controlled Vocabularies

Use of controlled vocabularies (keyword trees, taxonomies) leads to more consistent keywording and retrieval. It also helps refine the ambiguities of language and obtain more focussed results for the user.

- Preferred terms are the backbone of the keyword tree
- The structure organises the terms logically
- Terms can be defined unambiguously
- Associated synonyms assist the search
- Interoperability between systems is enhanced
- Translation between languages is enabled

IPTC controlled vocabulary field

The IPTC has defined a field (not yet published in 2010) to hold controlled vocabulary data in codified form.

The IPTC is considering the development of an IPTC Controlled vocabulary.

Finding the image

The word 'orange' can mean colour, fruit, or the mobile phone company.

- *If you search for blue flower you probably don't want red flowers with blue backgrounds.*
- *If you search for Paris, you might want either Paris the city, Paris Hilton, an event in Paris, or a typical image of Paris.*

9. IPTC Core and Extension

IPTC Core was created in 2004 using Adobe's XMP format for storing metadata in the image file. IPTC Core has the same fields as the legacy system IPTC-IIM, which is still used in some parts of the industry.

The Extension fields were added in 2008 to extend the Core schema and avoid ambiguities. Many of the rights fields added in Extension are adopted from the PLUS schema (PLUS License Data Format) developed in cooperation with the IPTC. See the [Interactive IPTC Field chart](#) for an overview of the IPTC Core and Extension fields.

There are 3 main categories of data:

Administrative – contact information for supplier, rights holders and licensors of the image

Descriptive – description of the visual content

Rights Information – copyright and underlying rights in the visual content including model and property rights, rights usage terms, licensing information

The numbers game

IPTC Core has 31 fields, and Extension has added more, to make a total of 77 fields. This can sound daunting, but here are some things to consider:

- *IPTC Extension widened the scope of the schema to include the needs of the stock industry and the cultural heritage sector.*
- *Templates make the capture of Exif data and the embedding of standard data easy and time effective.*
- *Not all fields are required for every business situation.*
- *The fields have been added where there was a need for more granular information, or where the use of the current fields is ambiguous (unclear semantics).*
- *The amount of data you require is up to you, but there are now more fields to choose from, with precise definitions.*
- *Use of IPTC Core and Extension should future proof your metadata and make it more interoperable.*

10. New fields in IPTC Extension

The fields added in IPTC Extension are listed on [IPTC Extension fields and how they are used](#) with some notes about their use.

PLUS fields

Some of the fields in IPTC Extension were adopted from the [PLUS](#) schema where they already existed. This provides interoperability between the IPTC and PLUS schemas.

New areas covered by Extension fields

Picture Number

Location

Model Information

People, organisations, events shown in the image

Artworks

Image Supplier information

Maximum size available, and Source File Type

Rights and licensing information

Go to [IPTC field definitions](#)

for the full definitions and the official user guide.

The PLUS Standards

The PLUS coalition has developed a set of standards for communicating and managing rights metadata. The PLUS License Data Format holds information on creators, licensors, models, underlying image rights, rights offered and rights granted under a licence.

The PLUS standards can be codified in machine readable form, allowing for automated workflows. PLUS has created a Matrix of media permissions a Glossary of rights terms, standardised rights packages and a system of identifiers for images, creators and rights.

a) Picture Number

The picture number, or Image ID number is the reference number that allows people to identify and refer to the image. In IPTC Core, there was no specific field for picture number, so the Title field is often used for this purpose.

New image ID fields in Extension

- ***Supplier's Image ID*** allows picture agencies to enter a new supplier ID without overwriting the Title field (which can still be used for photographer image number). This field is shared with PLUS and can link to a registry entry.
- ***GUID*** is a globally unique ID number for the image which should be created automatically at image creation in camera or scanner. This number retains a link to the creation of the image which doesn't change.

IPTC Core – improvements needed

- *The Title field can hold both text and numbers, and its semantics are not precise enough to identify the field uniquely as an image ID.*
- *There is no separation of the photographer picture number and the supplier (picture agency) picture number.*
- *The Title field information is commonly overwritten by agencies and clients, erasing the previous photographer or agency picture number.*



b) Location

Which location do we mean, the place where the photographer is standing, or the location of what's shown in the picture?

New fields in Extension

- **Location Image Created** – where the photographer was standing
- **Location Shown** – location(s) visible in the image.



IPTC Core – improvements needed

- *The IPTC Core Location field has been widely used as both location shown and location photographed. These can be different. E.g. shooting the Canadian part of the Niagara Falls from the US side.*
- *With GPS increasingly used to locate where the image was shot, it is important to create a separation.*
- *The IPTC created two new fields with clear semantics to avoid confusion with legacy images*

To note:

The original location fields in IPTC Core have been retained for legacy use, but the new fields should be used where possible.

The IPTC panels for use in Bridge allow users to copy information from Location Created and Location Shown fields to populate relevant fields.

11. Planning your metadata

Assess your business needs

Rights related fields

It is important to have rights data associated with images. You may want to make data entry in some fields mandatory before images are published to your database e.g. Copyright, Creator, Supplier's ID, Rights Usage Terms, Model rights,

Customer facing information

What data do your customers need?
e.g. Description, Supplier's Image ID, Credit, Headline

Internal documentation and administration

What data do you need for internal documentation and administration?
e.g. Job ID, Instructions, Creator contact details

Retrieval data

What data do you need to retrieve the media assets you are holding?
e.g. Keywords, Description

Workflow data

What data do you need to enable your workflow?
e.g. job or shoot number, legacy image numbers

Custom XMP fields

It is possible to create custom XMP fields to carry non IPTC data in the image file, but that will be for internal use as it is not immediately visible to outside users.

Can your customers read your data?

If your customers' software cannot read data in some IPTC fields, it can be merged, with suitable separators, in a widely used field such as Caption. (The data should be retained in the correct fields as well).

Non-IPTC Data

You need to take into account data which may not be catered for in the IPTC workflow. You and your customers may have additional fields which you can access via your software.

Assess your resources

- Where in the workflow is data most appropriately entered?
- Are these data entry points realistic for your business?
- Is training or communication needed by your staff and suppliers?
- Is your software helping or hindering your use of metadata?
- What scope do you have for changing or adding to your current software configuration



Be realistic, but look to the future

Many suppliers are very willing to enter basic data if they are given clear instructions. After all, the data may protect their own rights as well as easing their client's workflow.

Talk to your software suppliers

Your software suppliers may need to be informed about your metadata needs, and current best practice.

12. Use this guide to help you plan

IPTC Field Chart shows all the fields in IPTC Core and Extension. You can use it to help plan to plan your use of data.

1. Note the fields you need in your workflow.
2. Decide who should enter the data
3. Create stages for the supply chain
4. Assign a stage to the metadata fields you have selected
5. Review and pilot the workflow

Our sample Workflows are starting points for you to create your own metadata workflow.

We have examples for:

- **News agency**
- **Stock agency**
- **Museum**

Using the **interactive chart**

You can view the sample workflows and build your own custom scenario. Each workflow stage can be named, and assigned to fields on the chart to reflect the way you work. You can save your own workflow chart, and print it out for reference.

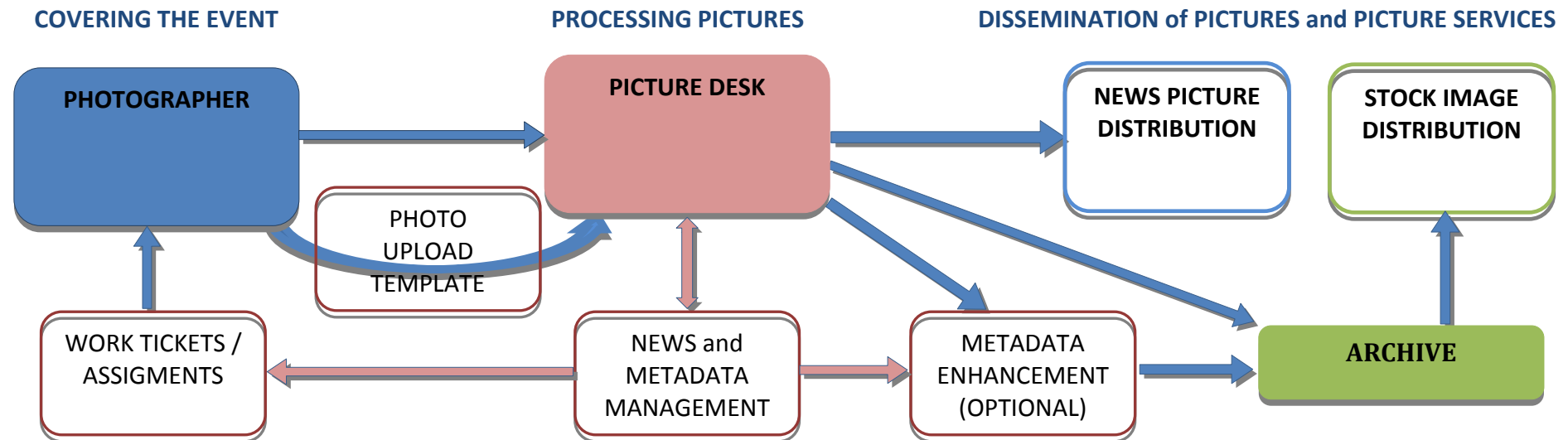
Every business is unique

The sample workflows show how a metadata workflow can be planned as a number of stages. Every business is unique in how it handles data, so these are starting points, and not prescriptive in any way.



13. News agency sample workflow

BUSINESS PROCESSES:



The Picture Desk manages coverage and distribution of news material. News agencies work under pressure of time so metadata may be forwarded to the photographer in work tickets and upload templates before images are ingested. At the picture desk metadata is edited and processed for news distribution and sent to the archive.

Photographer

The photographer enters data which identifies the shoot and the content of the image. This data may be in the form of a work ticket provided by the picture desk in advance of the shoot.

The Archive

Metadata may be enhanced for archive and stock image distribution, e.g. including keywords, supplier name and PLUS data.

IPTC Extension and agency software

Data entry will be enhanced when news agency software is updated to handle IPTC extension fields. Many agencies still use legacy IIM fields, which are limited in scope.

Managing Metadata

Picture desk management of news event coverage includes the allocation of metadata, even before the picture is taken. Metadata may be synchronized with news in text, audio or video or any other media type.

14. Stock agency sample workflow



Stock agencies want to receive as much relevant data from the photographer as possible, and this mostly relates to the specifics of the shoot. In our example, the agency does the keywording.

Photographer

Adds all relevant shoot data, with *Creator*, *Headline* and *Description* (the caption) as minimum. The *Instructions* field may give details of limitations on how the image can be used.

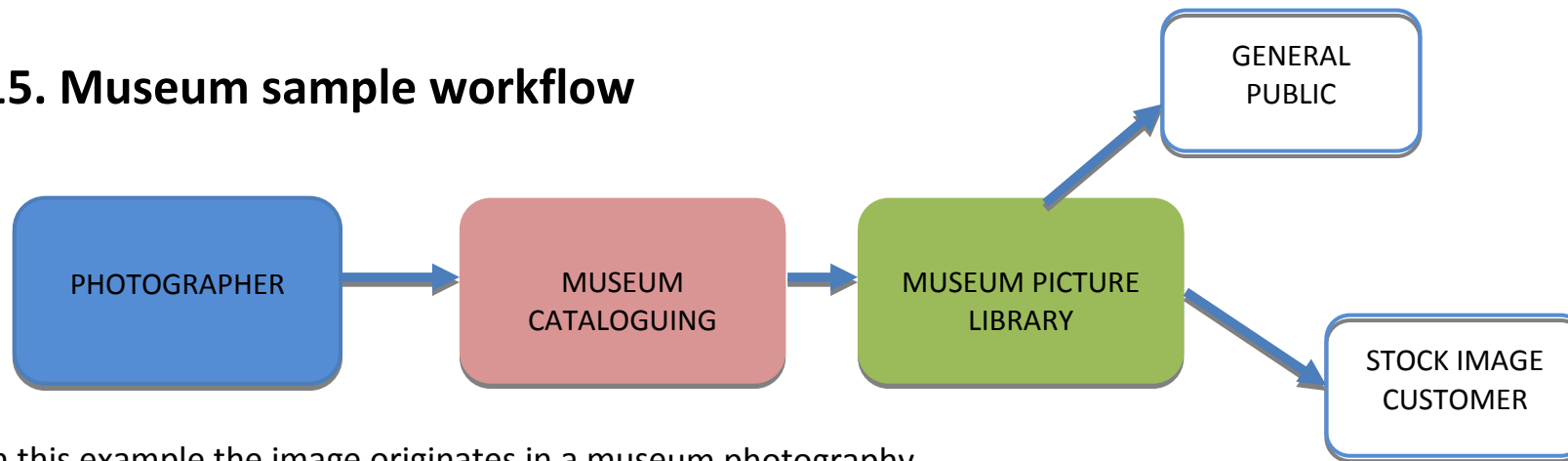
Stock Agency

Will add *Keywords*, *Credit line* and other rights data.

Photographer keywording

Most agencies ask photographers for some keywording. Some have systems for checking the incoming keywords against their in-house controlled vocabularies to enhance the data.

15. Museum sample workflow



In this example the image originates in a museum photography department, is catalogued by the museum, and entered into the museum content management system. It is then exported to the picture library where the metadata is enhanced for public distribution and/or image sales.

Photographer

Adds *Creator Name* and *Job Title* , *Date* and *Location Created*, along with *Job* or *shoot ID* and any *Instructions*.

Cataloguing/CMS system in Main Museum Departments

Cataloguing staff enter data about the artwork.

Picture Library

Adds *Keywords* and rights data needed for public access and sales of reproduction rights.

Museum web sites

Museums generally have a main web site where images can be viewed by the general public. This is separate from the picture library site (which may also be viewed by the public).

16. Resources

IPTC, Photo Metadata

www.iptc.org/photometadata

Standards for metadata embedded in the image file.

IPTC Field specification

www.iptc.org/site/Photo_Metadata/IPTC_Core_&_Extension/

Photometadata.org

www.photometadata.org

Information, tutorials and news at produced by SAA in partnership with Library of Congress.

Metadata Working group

www.metadataworkinggroup.com

Guidelines for handling image from leading companies Adobe, Canon, Microsoft, Nokia, Sony

PLUS

<http://www.useplus.com>

PLUS specification and Picture licensing Glossary



17. Acknowledgements

This guide has been written with the help of the IPTC Photo Metadata Working Group. Members of the group have checked accuracy, reviewed workflows, and contributed valuable on-the-ground experience.

With thanks to Michael Steidl IPTC Managing Director, David Riecks, Klaus Sprick, Cindy Lewis, Annette Feldman, Harald Loeffler, Andrea de Polo.

*Thanks to reviewers Andreas Gebhard (Getty Images) and Greg Reser (VRA).
Thanks to Alinari for the final design.*

Many thanks also to Graeme Cookson for creating the interactive PDF and the final PDF package.

Sarah Saunders, [Electric Lane](#), April 2011

*Published by CEPIC.
Fritschestraße 22
10585 Berlin
Germany
Tel. +49.30.88 91 01-60
www.cepik.org*

Photographs: copyright protected [Alinari](#) 24 ORE, Firenze, Italy

