

EXIF, IPTC, XMP ET AUTRES METADONNEES

De la théorie à la pratique

Par Jean-Paul Figer

Les métadonnées pour les photos.

- EXIF
- IPTC
- EXIF ou IPTC ?
- La convergence vers XMP

Le classement des photos.

- 1. Le nommage des photos.
- 2. La description des photos.
- 3. Les outils de recherche.
- 4. La fabrication des albums de photos ou de catalogues.

Annexes : Où mettre les métadonnées ?



Cet article est la suite de l'article [Le multimédia à la maison](#) dont la lecture préalable est recommandée.

La photo numérisée de bout en bout envahit votre ordinateur. Comment archiver, classer et retrouver facilement des milliers de photos dans votre ordinateur ? Il existe de nombreux programmes qui proposent des solutions toutes différentes. Laquelle choisir ? Comment être sûr de pouvoir réutiliser demain ce qui a été fait aujourd'hui ? Le classement des images est encore un sujet de Recherche. Il peut sembler hors de portée pour un amateur. Ce n'est heureusement plus le cas. Dès maintenant, vous pouvez obtenir de très bons résultats avec une panoplie de logiciels gratuits à condition de bien découper les tâches et de vous organiser en

conséquence.

Il y a 4 tâches principales :

1. Le stockage et le nommage des photos,
2. La description des photos,
3. Les outils pour la recherche,
4. La constitution des albums sur le PC ou pour une publication sur CD, DVD ou sur le Web

Avant de passer à la pratique, je vais dévoiler le plus simplement possible un des secrets les mieux gardés de l'Informatique : l'utilisation des métadonnées en particulier dans les photos numériques. Ensuite, j'expliquerai comment en tirer partie pour archiver, classer, retrouver et présenter vos photos sur votre ordinateur ou sur le Web.

LES METADONNEES POUR LES PHOTOS

En informatique, une métadonnée est une donnée qui décrit une donnée ou une ressource (document, image, fichier .mp3, page html, etc.). Quand vous faites clic droit sur un fichier puis propriétés dans Windows, le système vous présente les métadonnées de cette ressource, c'est à dire un ensemble d'informations sur ce fichier : type de document, taille, date de modification, sécurité, partage, etc.. Ce sont les métadonnées du système d'exploitation. Chaque type de document peut aussi contenir des métadonnées. Ce sont par exemple les balises ID3 des fichiers .mp3. Dans un document Word, faites Fichier, propriétés pour découvrir une foule d'informations sur ce document comme le nom présumé de l'auteur, les statistiques de frappe, le nom de la secrétaire, des mots-clés, des commentaires, la plupart non renseignées d'ailleurs. Ces métadonnées sont enregistrées et transmises avec le

document. Il est toujours amusant d'ailleurs quand on reçoit un document d'une entreprise qui se dit soucieuse de sécurité, de constater à quel point ces métadonnées sont des sources d'informations indiscrettes : document repris d'un fournisseur ou d'un concurrent, version précédente du document. L'utilisation de ces métadonnées reste aléatoire, car elles dépendent du système d'exploitation ou du produit et rien ne garantit qu'elles sont ou seront conservées lors des copies successives. Je déconseille d'ailleurs d'utiliser les métadonnées proposées par Microsoft dans Windows pour les photos car elles ne sont ni compatibles entre les différentes versions de Windows (?!), ni conformes aux standards!

A quoi servent ces métadonnées ? Essentiellement à identifier et retrouver la ressource. Depuis plus de deux mille ans, les bibliothécaires utilisaient des fiches ou des catalogues pour décrire et classer les documents de leurs bibliothèques. Ces fiches, plus faciles à manipuler que les documents eux-mêmes, ont donné naissance au concept de métadonnées. Ces fiches étaient produites par des humains pour des humains avec des manières très différentes. Pour rendre ces données utilisables facilement par les ordinateurs, il faut les standardiser. Il n'est donc pas étonnant que cette communauté ait été à l'origine d'une initiative visant à standardiser l'utilisation des métadonnées quel que soit le domaine d'utilisation. C'est le **Dublin Core Metadata Initiative ou DCMI**. Les plus curieux auront peut-être remarqué qu'à la fin de chaque article de mon site www.figer.com, en cliquant sur "métadonnées associées à cette page", on obtient un fichier RDF en XML qui décrit le document HTML conformément aux **15 éléments de base du DCMI**. Internet et le Web ont rendu ces métadonnées de plus en plus indispensables pour pouvoir effectuer des recherches plus intelligentes que sur de simples mot-clés. Si je tape mon nom "figer" dans Google, j'obtiens aussi toutes les occurrences du verbe "figer". Avec des métadonnées, si je recherche un auteur, je n'obtiens que toutes les occurrences de "figer" comme auteur. Quand les ressources sont des images, une recherche textuelle ne produit pas de résultat et les métadonnées s'imposent.

L'avantage du DCMI est de décrire de manière très formelle n'importe quelle ressource en précisant la valeur de ses propriétés. Cependant comme ses travaux sont récents, chaque communauté a défini ses propres métadonnées sans référence au DCMI. Le monde de l'image et de la photo a été particulièrement productif et il faut vivre aujourd'hui avec plusieurs types de métadonnées : **EXIF**, **IPTC**, **XMP**. Ce sont ces métadonnées que je vais décrire maintenant en indiquant au passage comment s'en servir. Toutes ces métadonnées sont intégrées dans le document JPEG en plus de l'image. Elles sont donc en permanence disponibles avec l'image et sont donc indépendantes d'un système ou d'un produit. **Voir ici les différentes manières de stocker les métadonnées.**

Pour faire des travaux pratiques à la lecture de cet article, je vous suggère d'utiliser le **logiciel de traitement d'image gratuit IrfanView**, déjà signalé dans mon article sur **Photos et images numériques**. Il faut l'installer avec tous ses plugins. Pour faire apparaître les métadonnées, il faut ouvrir une image puis taper I.

Vous pouvez aussi utiliser le meilleur logiciel gratuit que j'ai trouvé pour les métadonnées **PixVue** mais dont l'utilisation, entièrement intégrée dans l'explorateur de Windows, est un peu déroutante au début si on ne lit pas la notice d'utilisation.

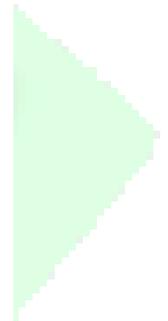
EXIF

EXIF est l'abréviation de "EXchangeable Image File". Ce format définit les informations techniques concernant les paramètres de prise de vue et les réglages de l'appareil photo au moment de la capture numérique. Le format EXIF a été développé en 1995 par le JEIDA (Japan Electronic Industry Development Association). La dernière version 2.1 date de Juin 1998 et est disponible sur le site **Exif.org**. Ces informations sont contenues dans le fichier image lui-même. Dans IrfanView, il suffit de cliquer sur le bouton "EXIF info" dans les propriétés de l'image pour les obtenir. Dans PixVue, il suffit de faire dans l'explorateur de fichiers, clic droit, propriétés. Voir ci-après à gauche un exemple extrait d'IrfanView et à droite un exemple produit par PixVue.

File: - D:\Photos\2004-07-04\2004-07-04-abbaye_des_vaux_de_c 003.jpg

Make - Canon
Model - Canon PowerShot A80
Orientation - Top left
XResolution - 180
YResolution - 180
ResolutionUnit - Inch
DateTime - 2004:07:03 19:16:42
YCbCrPositioning - Centered
ExifOffset - 196
ExposureTime - 1/100 seconds
FNumber - 4.90
ExifVersion - 0220
DateTimeOriginal - 2004:07:03 19:16:42
DateTimeDigitized - 2004:07:03 19:16:42
ComponentsConfiguration - YCbCr
CompressedBitsPerPixel - 3 (bits/pixel)
ShutterSpeedValue - 1/101 seconds
ApertureValue - F 4.91
ExposureBiasValue - 0.00
MaxApertureValue - F 4.91
MeteringMode - Multi-segment
Flash - Not fired, auto mode
FocalLength - 23.41 mm
UserComment -
FlashPixVersion - 0100
ColorSpace - sRGB
ExifImageWidth - 2272
ExifImageHeight - 1704
InteroperabilityOffset - 1844
FocalPlaneXResolution - 8114.29
FocalPlaneYResolution - 8114.29
FocalPlaneResolutionUnit - Inch
SensingMethod - One-chip color area sensor
FileSource - DSC - Digital still camera
CustomRendered - Normal process
ExposureMode - Auto
WhiteBalance - Auto
DigitalZoomRatio - 1.00 x
SceneCaptureType - Standard

Maker Note (Vendor): -
Macro mode - Normal
Self timer - Off
Quality - Fine
Flash mode - Auto + red-eye reduction
Sequence mode - Single or Timer
Focus mode - Single
Image size - Large
Easy shooting mode - Full Auto
Digital zoom - None
Contrast - Normal
Saturation - Normal
Sharpness - Normal
ISO Value - Auto
Metering mode - Evaluative
Focus type - Auto
AF point selected -
Exposure mode - Easy shooting
Focal length - 250 - 749 mm (32 mm)
Flash activity - Not fired
Flash details -
Focus mode 2 - Single
White Balance - Auto
Sequence number - 0
Flash bias - 0.00 EV

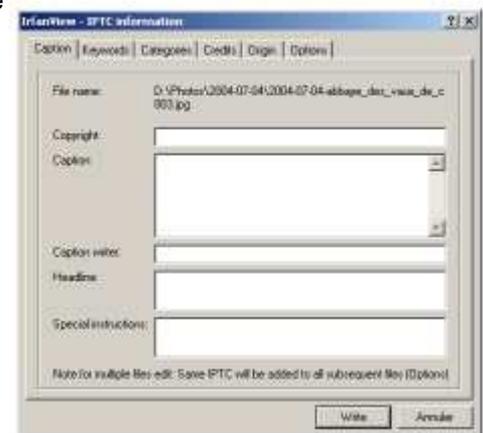


Subject Distance - 408 mm
Image Type - IMG:PowerShot A80 JPEG
Firmware Version - Firmware Version 1.00
Image Number - 1020219
Owner Name -

Ces données sont produites à la prise de vue et sont de type interne c'est à dire qu'elles sont intégrées dans l'image. Elles sont transportées avec l'image. En principe, elle ne doivent pas être modifiées. Cependant le logiciel gratuit **Photo Studio** vous permet de les modifier si nécessaire, en particulier d'ajouter la position GPS, le copyright ou autre. Il permet aussi de voir les commentaires JPEG et la vignette que mon appareil photo insère dans l'image.

IPTC

L'IPTC -International Press and Telecommunications Council- est une organisation internationale créée en 1965 pour développer et promouvoir des standards d'échange de données à destination de la presse. En association avec la NAA -Newspaper Association of America-, l'IPTC a défini un modèle global de données appelé "IPTC-NAA Information Interchange Model" connu sous le nom IIM. Un sous-ensemble de ce modèle a servi de base à la société Adobe pour définir dans son logiciel Photoshop les informations associées à une image. C'est ce sous-ensemble qui est communément appelé métadonnées ou champs IPTC. Ces métadonnées sont de type interne, c'est-à-dire qu'elles sont stockées à l'intérieur des fichiers images JPEG. Elles voyagent donc avec l'image. Pour ajouter ces métadonnées, il suffit de cliquer sur File info... dans Photoshop, dans la page de propriétés de l'image sur "IPTC info" dans IrfanView ou clic droit sur le fichier, annotations, éditer dans PixVue.



Voici un **tableau de correspondance entre Adobe Photoshop et IrfanView** pour le nommage des champs IPTC.

EXIF ou IPTC ?

On constate qu'il existe de nombreux champs en double entre les métadonnées EXIF et IPTC ce qui est inévitable puisque les 2 normes ont été produites séparément. Une bonne règle consiste à ne pas toucher aux données EXIF qui sont produites automatiquement à la prise de vue et à renseigner à la main les métadonnées IPTC qui concernent le contenu de l'image (créateur, description, copyright,...). Tous les logiciels cités permettent de faire des traitements par lots pour ajouter le même champ à un ensemble d'images. Je vous conseille cependant de faire une sauvegarde préalable de vos photos. Ces opérations modifient le contenu des images en remplaçant l'image d'origine.

Le nombre de champs IPTC est très élevé. Ne remplissez que ceux qui vous seront utiles pour des classements, légendes ou recherches.

La convergence vers XMP

XMP a été créée par Adobe en 2001. Il utilise un format XML et s'appuie sur RDF et sur les éléments du Dublin Core. Il définit aussi une manière, XMP Packet, d'encapsuler les données dans les applications. Il n'est donc plus limité au format JPEG mais est aussi défini pour TIFF, GIF, HTML, PDF, XML/SVG. Il est de plus extensible. Une vraie merveille disponible dans les dernières versions des produits Adobe. Ce format n'est pas encore très répandu mais ce ne saurait tarder puisqu'en juin 2004, l'IPTC et Adobe ont annoncé qu'ils travaillaient ensemble pour intégrer IPTC et XMP. Il existe un logiciel gratuit très intéressant **PixVue** qui s'intègre dans l'explorateur de fichiers Windows et qui permet de voir et mettre à jour les métadonnées XMP, IPTC et EXIF. Il permet aussi d'extraire ou d'exploiter ces métadonnées. Comme il met à jour simultanément les champs XMP et IPTC, vous n'avez pas de souci de transposition. C'est le logiciel que je recommande pour ce travail bien que l'interface utilisateur ne soit pas grand public.

☐ LE CLASSEMENT DES PHOTOS

Il n'y a pas de méthode toute faite, et chacun doit définir sa manière de faire. Néanmoins, la disponibilité pour pas cher de disques de grande capacité et d'outils de recherche performants permet de simplifier le travail en "laissant faire" l'ordinateur. La stratégie générale consiste à **ne pas classer** et à tout mettre en vrac dans un dossier. Pour des raisons pratiques, on peut faire des sous-répertoires par date. Ensuite, si on a le temps, on ajoute des métadonnées IPTC/XMP qui décrivent les photos pour pouvoir faire des recherches ou pour produire automatiquement des albums avec des descriptions qui accompagnent les photos. Un outil comme **Picasa** permet d'examiner ces dossiers de manière entièrement automatique. Je conseille son usage pour cette visualisation intuitive. La version 1 de ce logiciel n'exploitait pas les métadonnées IPTC et EXIF, ce qui était quand même un comble quand on venait d'être racheté par Google. Cette lacune flagrante est en partie comblée dans la version 2 sortie en janvier 2005 puisque les mot-clés et les descriptions sont enfin stockées sous forme de métadonnées dans les photos ! Pour constituer des albums et des diaporamas sur votre PC, vous pouvez utiliser Picasa car la simplicité d'utilisation est exceptionnelle. Picasa propose aussi le logiciel Hello qui permet de télécharger automatiquement les photos sur un blog de **Blogger.com**. Si c'est pour faire un album sur le Web, **JAlbum** qui exploite les métadonnées IPTC est la meilleure solution ou **PixVue** qui permet de tout faire mais avec un peu plus de travail. En fait tous ces logiciels constituent des albums de vignettes qui pointeront sur les photos en définition originale ou qui les recopieront dans d'autres dossiers. La taille des vignettes est de quelques Ko ce qui permet des affichages simultanés de photos très rapidement. Autre avantage, la même photo peut se retrouver dans plusieurs albums à la fois. On peut aussi détruire et refaire des albums sans toucher aux images d'origines. Ceci permet de préparer des documents à publier sur le web ou sur CD et DVD dans différents formats. On peut publier les métadonnées dans les albums ou dans des pages web spéciales. Avec un outil de recherche sur le PC comme Google Desktop, on retrouve alors instantanément dans son ordinateur toutes les photos qui concernent un sujet, une catégorie ou un nom. Enfin, la puissance de l'ordinateur est à votre service! Voici donc comment procéder, sachant que toutes ces étapes, sauf la copie des photos dans un dossier de votre PC, sont optionnelles ou réalisables quand on a le temps.

-1. Le nommage des photos.

Dans ce domaine, il faut rester très simple. En 2004 pour 100 euros, un disque de 200Go fournit un espace suffisant pour stocker 200 000 photos en format JPEG. Je conseille donc tout simplement de stocker les photos tel quel sur le disque dans un ensemble de répertoires et de sous-répertoires. On peut ainsi faire facilement une sauvegarde sur un DVD lisible sur les lecteurs DVD de salon modernes. Le nom de fichier est une métadonnée associée à la photo lisible par les programmes et les humains. Une bonne stratégie consiste à s'en servir comme outil de gestion simple d'une collection de photos. J'utilise comme identifiant majeur la date en **format ISO 8601** utilisé aussi en XML. Le format est AAAA-MM-JJ par exemple 2004-12-31. Le principal avantage de ce format est d'être classé correctement par un tri alphabétique. Ensuite, j'ajoute le lieu et un numéro d'ordre, ce qui donne un nom complet du genre 2004-12-31_Paris_123.jpg. J'utilise uniquement les caractères suivants a à z A à Z 0 à 9 \$-_.+! pour ne pas avoir de problèmes lorsque ces noms sont présentés dans des pages Web ou dans des noms de dossiers. En particulier, je n'utilise ni caractères accentués ni espace(_ à la place). Les photos peuvent au choix être placées en vrac dans un répertoire ou classées dans des sous-répertoires dont le nom est constitué par les premiers caractères des noms de fichiers qu'il contient. Ce travail est très facile à réaliser car tous les outils que j'ai cités permettent relativement simplement de faire des renommages automatiques et globaux de fichiers. je suggère quand même de faire des tests sur des copies de dossiers car une erreur "globale" est vite faite.

-2. La description des photos.

Il faut maintenant ajouter à chaque photo une description pour pouvoir faire des recherches ou obtenir automatiquement des légendes dans la publication des photos. Il faut bien sûr intégrer ces métadonnées dans les fichiers JPEG pour les transporter avec les images. Les métadonnées EXIF restent inchangées. Il faut ajouter les métadonnées IPTC et/ou XMP à l'aide de Photoshop, d'IrfanView ou de PixVue.

-3. Les outils de recherche.

Les outils de recherche se classent en 2 catégories. D'un côté, ceux qui construisent un catalogue de toutes les images disponibles et qui permettent de naviguer rapidement à l'intérieur du PC comme ACDSee, ZoomBrowser ou Picasa. La recherche est alors purement visuelle. C'est parfait pour quelques centaines de photos et pour quelqu'un qui se souvient vaguement du contenu. L'utilisation est immédiate. De l'autre, les systèmes qui utilisent les métadonnées. Ces systèmes permettent des recherches beaucoup plus rapides et beaucoup plus faciles quel que soit le nombre de photos. Ils permettent de constituer des bibliothèques en provenance de sources différentes et de rechercher sans une connaissance au préalable du contenu. C'est la voie à suivre depuis que des standards mondiaux émergent. C'est un vrai progrès à condition de faire l'effort de renseigner les métadonnées.

-4. La fabrication des albums de photos ou de catalogues.

J'utilise Picasa uniquement pour examiner l'ensemble du contenu du PC. Je fabrique systématiquement des albums ou des catalogues au format WEB(HTML) avec JAlbum. Ceux-ci peuvent donc être utilisés en local ou publiés tel quel sur le web ce qui est très pratique. Avec JAlbum, on dispose de 3 niveaux de définition ajustables : Le niveau vignette (100 à 200 pixels de côté, moins de 10 Ko), le niveau affichage web (600x450 pixels, environ 100Ko) et la définition d'origine que l'on peut télécharger (environ 1 Mo). On peut aussi faire une copie de ces albums sur CD ou DVD pour les lire sur un téléviseur avec un lecteur DVD de salon.

OU METTRE LES METADONNEES ?

Les métadonnées sont souvent stockées à l'extérieur des documents comme les fiches qui décrivent les livres d'une bibliothèque ou les propriétés qui décrivent les caractéristiques d'un fichier. Cette pratique permet de constituer des "bases de données" que l'on interroge sans référence aux documents eux-mêmes. Cependant, quand le document est sous forme numérique et qu'il est envoyé d'un système vers un autre, vous perdez les métadonnées associées sauf si vous les transmettez à part. Il est donc plus judicieux d'inclure ces métadonnées dans le document lui-même en les identifiant à l'aide de balises. C'est ce qui est fait, par exemple, pour les balises ID3 des fichiers musicaux MP3 ou pour les balises "<meta.. />" des pages HTML.

Cette manière de faire est donc bien meilleure si plusieurs conditions sont respectées :

- le contenu des balises est indépendant d'un système ou d'un langage particulier. C'est tout l'intérêt d'utiliser des données en XML.
- les différents logiciels de manipulation des documents "préservent" ces balises. Je vous suggère de vérifier que c'est bien le cas pour les logiciels de manipulation de photos que vous utilisez.
- les programmes de recherche savent extraire automatiquement les métadonnées de vos documents lors du stockage dans un système particulier.

A titre d'exemple, j'ai inclus des métadonnées dans l'image qui illustre cet article. Faites clic droit, enregistrer l'image sous... pour récupérer cette image sur votre PC et vous pourrez lire les métadonnées à l'aide des logiciels que j'ai indiqués.

