

LES METADADES DE LES IMATGES FIXES I ELS MITJANS SOCIALS

Patrick Peccatte

École des Hautes Études en Sciences Sociales, EHESS (París)

Resum

Després d'un petit recordatori dels diferents estàndards de metadades integrades en les imatges numèriques fixes, examinarem el comportament que adopten unes determinades plataformes socials en relació amb aquestes informacions. Per acabar, estudiarem les (escasses) possibilitats d'explotació de les metadades en les xarxes socials mitjançant l'experiència del projecte *PhotosNormandie* de la plataforma Flickr.

1. Recordatori dels diferents estàndards de metadades integrades en les imatges

Per tal de comprendre els dos procediments que permeten descriure imatges numèriques fixes amb l'ajuda de metadades, recordem els àlbums de fotos antics.

Per descriure una foto, per recordar-ne les dades, els llocs, les persones que hi apareixen, es fan unes llegendes breus al mateix àlbum, al costat de la foto. Ara bé, si les fotos es treuen de l'àlbum d'una manera o d'una altra i es desendrecen completament, es perd l'associació de les llegendes amb les imatges i pot ser difícil restablir aquest vincle.

Evidentment, tothom sap quina és la solució a aquest problema. N'hi ha prou d'escriure la llegenda al dors de la foto. D'aquesta manera la llegenda sí que queda unida a la foto.

Amb l'escriptura d'una llegenda en un àlbum o d'algunes paraules indicatives darrere d'una foto, els dos procediments essencials que permeten descriure la imatge fotogràfica queden representats des de fa molt de temps. Retrobem, en efecte, els mateixos principis amb els nostres objectes numèrics, traslladats al llenguatge modern; reconeixem, d'una banda, l'assignació de metadades *externes* a la imatge [la llegenda en un catàleg o en un àlbum] i, de l'altra, l'addició de metadades *internes* a la imatge, és a dir, la inserció de metadades integrades en el fitxer numèric que és la imatge [l'escriptura al dors de la foto].

Tot seguit ens centrarem en tres tècniques específiques de metadades d'imatges numèriques fixes:

RESUMEN

Tras un sucinto esquema sobre los diferentes standards de metadatos incrustados en las imágenes fijas (IPTC/IIM, Exif, XMP) la presentación examinará cómo se comportan ciertas plataformas sociales (Dropbox, Facebook, Flickr, Instagram, Pinterest, Tumblr, Twitter...) con dichos metadatos. En efecto, la mayoría de ellas ignoran los metadatos incrustados y los suprimen totalmente, lo cual es inaceptable tanto desde el punto de vista documental como en lo que concierne a los derechos asociados a las imágenes. La intervención finalizará con las (escasas) posibilidades de explotación de los metadatos en las redes sociales, particularmente a través de la experiencia del proyecto Photos Normandie que consiste en mejorar los títulos o leyendas de un fondo de fotos históricas basándose en los metadatos incrustados de las imágenes.

SUMMARY

After a brief outline of the different metadata standards embedded in still images (IPTC/IIM, Exif, XMP), this presentation will examine how certain social platforms (Dropbox, Facebook, Flickr, Instagram, Pinterest, Tumblr, Twitter...) behave with such metadata. Indeed, most of these platforms ignore embedded metadata, which are deleted completely, a fact that is unacceptable from a documentary point of view and also from what concerns the rights associated with images. This speech will conclude with the (limited) possibilities of exploitation of metadata in social networks; particularly through the experience of the project Photos Normandie, which aims to improve the titles or captions of historical photo collections from the metadata embedded in images.

EXIF
IPTC/IIM
XMP

EXIF és una abreviatura d'*EXchangeable Image File*.

El format EXIF va ser ideat a l'octubre del 1995 pel consorci japonès JEIDA (Japan Electronic Industry Development Association), que agrupava les indústries electròniques anteriors a l'any 2000. No és un estàndard, però el compleixen tots els fabricants d'aparells fotogràfics numèrics (APN), si bé n'hi ha variants propietàries.

Aquest format defineix els paràmetres de presa de vista i les configuracions de l'aparell en el moment de la captura numèrica.

Es tracta de metadades de tipus *intern*, entre les quals cal esmentar el nom del fabricant i el model de l'aparell, l'alçada i llargada de la imatge, la data i l'hora de la presa de vista, l'orientació, la resolució, el temps d'exposició, l'obertura, la presència o no d'un flaix, les coordenades GPS, etc.

L'IPTC (International Press and Telecommunications Council) és un consorci que reuneix les principals agències de premsa del món.

L'IPTC comprèn actualment una cinquantena d'organismes (entre els quals l'AFP i l'agència EFE, la principal agència de premsa en llengua espanyola al món).

L'IPTC desenvolupa normes tècniques d'intercanvi de dades per a la premsa.

Aquests estàndards són utilitzats per la pràctica totalitat de les agències de premsa del món.

L'IPTC i la NAA (Newspaper Association of America) van crear el 1991 el model global de dades denominat Information Interchange Model [IIM].

Des del 1994, la societat Adobe ha utilitzat un subconjunt simplificat de l'IIM per definir dins del seu programari Photoshop les informacions associades amb una imatge.

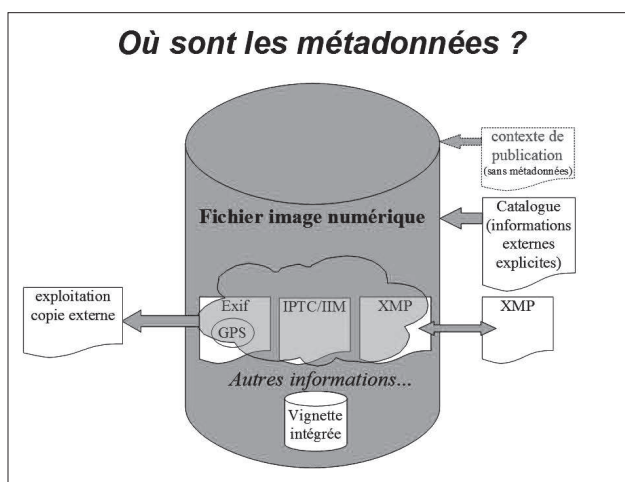
Les metadades IPTC/IIM són metadades de tipus *intern*. Es tracta d'un conjunt de camps textuais emmagatzemats en el fitxer imatge: títol, llegenda, mots clau, copyright, etc.

Aquest estàndard encara és molt utilitzat en la premsa i l'edició, tot i que és considerat obsolet i està sent substituït per XMP.

XMP (eXtensible Metadata Platform) va ser creat per la societat Adobe al setembre del 2001.

Aquesta tècnica usa una versió simplificada de RDF (Resource Description Framework), que és un estàndard desenvolupat per W3C, base del web semàntic; RDF permet codificar, intercanviar i reutilitzar metadades estructurades i es pot expressar en XML.

Les imatges proveïdes de metadades XMP constitueixen probablement la col·lecció d'objectes més



Imatge 1 - On són les metadades? Esquema dels estàndards de metadades d'imatges numèriques fixes.

voluminosa descrits en RDF al web, però la tècnica no està reservada a les imatges...

XMP utilitza l'esquema Dublin Core com a base.

És un estàndard ISO des del març del 2012 (ISO 16684-1, 2012). Ja no és, doncs, una tecnologia específicament Adobe.

En resum, les informacions EXIF (i el seu subconjunt GPS) i les informacions IPTC/IIM de les imatges numèriques fixes són metadades internes. Quant a les metadades XMP, tant poden ser internes com externes. Una observació: les metadades no adopten sempre una forma textual. Molt sovint, per exemple, una imatge pot contenir una vinyeta, una imatge de baixa resolució, que es pot considerar com una metadada. Més endavant tornarem a tractar el que aquí es denomina *context de publicació*.

Les metadades internes presenten un gran nombre d'avantatges. Mitjançant l'ús d'aquesta tècnica, es facilita l'intercanvi, ja que el recurs numèric transporta amb si mateix les seves pròpies metadades quan es descarrega, es copia, adopta un altre nom, es compacta, etc.

Ara bé, les metadades internes presenten també alguns inconvenients: cal extreure les informacions i copiar-les en una base de dades per explotar una gran col·lecció de recursos numèrics. De la mateixa manera que amb una llegenda inscrita al dors d'una foto l'accés a les informacions no és directe, l'usuari ha d'efectuar una operació per tenir coneixement de la llegenda.

Tornem al context de la *publicació*.

Google Images indexa solament el nom del fitxer imatge i el text que emmarca la imatge en la pàgina on apareix; és el que aquí rep el nom de *context de publicació* d'una imatge.

Per a un motor de cerca generalista, en efecte, les imatges que contenen metadades constitueixen una part insignificant de les imatges del web.

Google Images no indexa, doncs, les metadades internes de les imatges (IPTC/IIM o XMP).

Per comprovar-ho, n'hi ha prou d'efectuar una prova indexant una imatge amb un mot clau, *hàpax*, és a dir, un pseudomot clau únic, tant en IPTC/IIM com en XMP.

Ens podem preguntar, malgrat tot, si la presència de metadades internes en les imatges millora el posicionament en els resultats de recerca de Google Images. Es tracta aquí d'un interrogant recurrent en SEO (*Search Engine Optimization*, o optimització de motors de cerca).

Les proves efectuades en diferents períodes evidencien que *no*: les metadades no milloren el posicionament de les imatges en una cerca.

Ara bé, Google probablement té projectes interns que exploten les metadades internes en les imatges (no disposem d'aquesta informació).

2. Els mitjans socials i les metadades internes en les imatges fixes

El manifest *Embedded Metadata* (metadades incloses/integrades) de l'IPTC PhotoMetadata Working Group defineix cinc principis guia per a la creació i l'emmagatzematge de metadades, a fi que siguin traslladades amb el fitxer cada cop que sigui possible.

El manifest sosté que les metadades associades amb una imatge s'han de mantenir en totes les etapes del flux de dades (*workflow*), incloent-hi qualsevol exhibició pública.

El manifest s'adreça a tots els organismes que afegixen i administren les metadades, i també als proveïdors de materials i programaris dels sistemes dels quals exploten el flux de dades.

El manifest ha dut a terme recentment una enquesta pública titulada *How Social Media sites manage metadata?*

La metodologia i els resultats de l'enquesta són descrits en el lloc del manifest a l'adreça següent:

<<http://www.embeddedmetadata.org/social-media-test-results.php>>

El mètode utilitzat en l'enquesta és simple. Una imatge de prova que tingui un conjunt complet de metadades (EXIF, IPTC/IIM, XMP) es descarrega en diferents xarxes socials. A continuació observem les metadades d'aquestes imatges emmagatzemades en aquestes plataformes. El detall del protocol es descriu al lloc del manifest a l'adreça següent: <<http://www.embeddedmetadata.org/social-media-test-procedure.php>>

Es van efectuar dues sèries de proves el 2013 i més tard, el 2015, a fi de tenir en compte les noves xarxes socials i d'observar les evolucions eventuais entre les dues dates.

Els dos objectius principals eren analitzar precisament quines són les metadades integrades que apareixen en cada plataforma de xarxa social, i verificar les metadades conservades i les que són modificades o eliminades.

Es van definir quatre proves concretes:

Quines són les metadades incloses que es visualitzen en la interfície d'usuari?

Les informacions relatives a la targeta de crèdit es visualitzen correctament? (verificació de les «4C»: *caption, creator, copyright notice, creditline*)

Quines són les metadades que es conserven quan es recupera la imatge des d'un navegador, amb l'ajuda d'una ordre del tipus *Save As [Desa com a]*?

Quines són les metadades que es conserven quan la xarxa social proposa una descàrrega de la imatge (amb l'ajuda d'un botó de descàrrega, per exemple)?

Els resultats no són gaire brillants, certament.

Les xarxes socials més conegudes modifiquen les metadades integrades d'una manera o d'una altra. Fins i tot es pot valorar que la situació es deteriora a escala global, ja que els resultats van ser una mica millors el 2013.

Examinem ara alguns resultats per a diferents plataformes conegudes.

Dropbox

No es visualitza cap metadada.

Les metadades es conserven únicament amb una descàrrega, no es conserven fent un «Desa com a».

Deteriorament: el 2013, en canvi, sí que es conservaven fent un «Desa com a».

Facebook

No es visualitza cap metadada.

Únicament es conserven les metadades *Copyright Notice* i *Creator* de l'IPTC/IIM fent un «Desa com a». Totes les altres són eliminades.

Lleugera millora des del 2013: totes les metadades es van eliminar fent un «Desa com a».

Es constata també a Facebook una curiositat intrigant. La plataforma afegix sistemàticament dues metadades IPTC/IIM als camps *Special Instructions* i *Original Transmission Reference*.

És summament difícil entendre a què corresponen aquests codis generats en la descàrrega d'una imatge de Facebook. En efecte, no existeix cap comunicació de la societat sobre aquest tema i l'enginyeria inversa no és capaç en aquest cas de determinar el significat d'aquestes informacions. D'altra banda, l'IPTC ignora tota aquesta particularitat de les imatges que hagin passat per Facebook.

Gràcies a aquest dispositiu, Facebook potser és capaç d'efectuar un seguiment bàsic de les imatges que han transitat per la plataforma.

Flickr

Algunes metadades es visualitzen correctament, però no totes les «4C».

Totes les metadades es conserven després de fer una descàrrega o un «Desa com a» de la imatge en la seva definició original, però són eliminades en les altres definicions.

Deteriorament: cap al 2010, totes les resolucions intermèdies proposades per la plataforma contenen les metadades de la imatge original.

Google Photos

Algunes metadades es visualitzen correctament, però no totes les «4C».

Les metadades es conserven després d'una descàrrega de la imatge original.

Només les metadades Exif es conserven fent un «Desa com a» en les imatges en baixa resolució.

Deteriorament: el 2013, totes les metadades es conservaven fent un «Desa com a» en les imatges en baixa resolució.

Instagram

No es visualitza cap metadada.

No es pot fer cap desament.

El 2013, es podia fer un «Desa com a», però així s'eliminaven les metadades.

Instagram és una de les pitjors xarxes socials pel que fa a les metadades internes de les imatges fixes.

Pinterest

No es visualitza cap metadada.

Les metadades es conserven fent un «Desa com a» de la imatge en la seva definició original, però no amb una descàrrega.

No es va provar el 2013.

Tumblr

No es visualitza cap metadada.

Només les metadades Exif es conserven fent un «Desa com a», totes les altres s'eliminen.

Deteriorament: el 2013, totes les metadades integrades es van conservar fent un «Desa com a».

Twitter

No es visualitza cap metadada.

Només les imatges en baixa resolució estan disponibles fent un «Desa com a», sense cap metadada.

No s'ha modificat des del 2013.

Twitter se situa a la cua, juntament amb Instagram.

Aquests mals resultats no són una fatalitat. Sens dubte, és possible concebre les xarxes socials respectuoses amb metadades integrades, com fa Behance, plataforma pertanyent a Adobe que aplega dossiers de presentació d'artistes (enllaç www.behance.net).

Behance

Totes les metadades 4C es visualitzen correctament.

Diverses altres metadades (però no totes) es visualitzen igualment.

Totes les metadades es conserven amb una descàrrega i fent un «Desa com a».

3. Treballar amb metadades internes en un mitjà social, el projecte PhotosNormandie de la plataforma Flickr

Tot i aquests resultats poc encoratjadors, es pot treballar de manera eficaç en una xarxa social «mediocre» amb les metadades internes de les imatges.

El projecte *PhotosNormandie* de Flickr té com a objectiu millorar la descripció documental d'un fons de més de 3.400 fotografies històriques de la batalla de Normandia (des del desembarcament del 6 de juny de 1944 fins al final d'agost de 1944).

És un projecte de *crowdsourcing* (o proveïment participatiu, que són continguts generats pels usuaris), actiu en la plataforma multimèdia Flickr des del gener de 2007 a l'adreça següent:

www.flickr.com/photos/photosnormandie/.

És obert a tothom. Un total d'una seixantena de col·laboradors han treballat en *PhotosNormandie* i una desena de parts interessades participen de manera regular en el projecte.

Quin és l'origen de les fotos del projecte *PhotosNormandie*?

2.760 fotos provenen del lloc *Archives Normandie 1939-1945*, que ja no existeix avui; era un servei públic del Conseil Régional de Basse-Normandie, creat el 2004; totes aquestes fotos provenen dels Arxius Nacionals dels Estats Units i del Canadà i estan «lliures de drets» (en el sentit nord-americà).

296 fotos de *The Allison Collection*; són fotos retransmeses per ràdio el 1944.

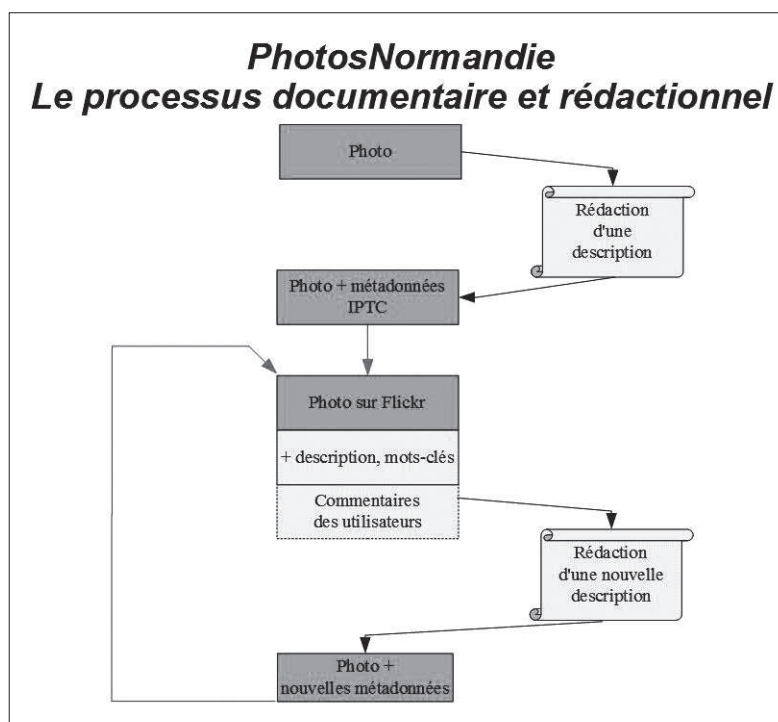
322 fotos procedeixen de la biblioteca de la ciutat de Cherbourg-Octeville.

163 fotos procedeixen de la mediateca de Lisieux.

Les llegendes de les fotos estan escrites segons els estàndards de metadades IPTC/IIM i XMP.

Diversos camps textuais bàsics (títol, llegenda, mots clau, copyright, etc.) s'emmagatzemen en el fitxer imatge. El mètode parteix d'una funció poc comuna de la plataforma Flickr: la compleció automàtica de camps Flickr a partir de camps IPTC quan es descarrega una foto.

La utilització sistemàtica dels estàndards IPTC/IIM i XMP en el projecte *PhotosNormandie* concreta els avantatges de les metadades internes. La descripció



Imatge 2 - Descripció del procés documental i de redacció en el projecte PhotosNormandie

textual de la imatge sempre està disponible amb la imatge i es pot reutilitzar fàcilment. L'usuari queda alliberat de la tecnologia de bases de dades utilitzada per a l'explotació del seu corpus d'imatges.

Es pot resumir aquest punt de vista així: les metadades són com les imatges, us pertanyen, no pertanyen al vostre prestador de serveis. En el cas del projecte *PhotosNormandie*, per exemple, podem sortir fàcilment de Flickr i seguir el projecte en una altra plataforma que sigui compatible amb un conjunt reduït de metadades integrades (com 500px, fotki, ipernity [clon de Flickr], joomeo, smugmug).

El mètode utilitzat a Flickr presenta, malgrat tot, alguns inconvenients. En efecte, la posada al dia d'una descripció és feixuga, ja que cal tornar a descarregar cada foto que contingui una nova descripció. Però sobretot una URL Flickr que indiqui una foto no es pot considerar estable perquè el número d'identificació Flickr canvia quan tornem a descarregar la imatge. No obstant això, l'accés directe a cada foto continua sent possible, ja que es pot treballar amb les referències (ID) de les fotos, i els avantatges del mètode compensen en gran mesura els inconvenients que té.

El balanç preliminar del projecte *PhotosNormandie* és extremament positiu: des del final de gener de

2007, la galeria i les fotos van ser vistes més de 36 milions de vegades (és a dir, més de 7.700 visites cada dia; el 6 de juny de 2014, es van comptabilitzar prop de 200.000 visites).

Es constata un gran progrés des de fa uns quatre anys (4.500 visites al dia el 2012).

9.222 descripcions han estat corregides i actualitzades (algunes descripcions s'han corregit diverses vegades).

417 fotos corresponen a seqüències gravades recuperades (més d'1 de 10, i encara n'hi deu haver més).

El text en francès d'aquesta ponència es pot consultar a:
http://www.girona.cat/sgdap/cat/jornades_actes.php