Metadades en fitxers

Per "metadada" s'entén tota aquella informació que acompanya, explica i posa en context a qualsevol dada informàtica. Per exemple, en el cas d'imatges, metadades que poden estar incrustades dins del propi fitxer de fotografia són la geolocalització -latitud i longitud- des d'on es va prendre la fotografia (si aquesta metadada la genera la càmara, no totes ho fan), la configuració de la càmara (la velocitat ISO, la d'obturador, tipus de lents, etc) i la seva marca i model, la data i hora quan es va prendre la fotografia, el format de la imatge, el nom del programa utilitzat per editar la fotografia (si això s'escau), el "copyright" (si això s'escau), etc. En el cas de documents PDF, exemples de metadades incloses dins del propi document són el nom de l'autor/usuari, el nom i versió del programa utilitzat per crear-lo, el títol del document, la data i hora de creació i/o d'última modificació, etc. En el cas de processadors de textos, a més de les metadades anteriors, també s'hi podrien afegir línies de text i comentaris sencers esborrats en versions prèvies del document (¡atenció!). En el cas de vídeos, ens podem trobar amb metadades que són automàticament generades (duració, mida, format, còdecs emprats, data de creació, geolocalització -si s'escau-,...) i les que són afegides manualment (transcripcions de text, etiquetes, notes per editors, etc). En el cas de fitxers d'àudio, les metadades són similars a les comentades dels fitxes de vídeo però, a més, existeixen diferents formats específics de metadades per emmagatzemar informació específica com informació de l'artista, gènere musical, número de cançó en un àlbum, miniatura de la portada de l'àlbum, etc.

Existeixen diferents eines que ens permeten observar -i segons les circumstàncies, fins i tot editar i/o esborrar!- les metadades mencionades al paràgraf anterior; dependrà del tipus de fitxer (si és una imatge, un PDF, etc) es faran servir unes eines o unes altres; als propers exercicis en veurem exemples diversos. Tingueu en compte que esborrar metadades d'un fitxer és una acció bàsica de seguretat per mantenir l'anonimat

Més enllà d'aquestes metadades particulars (que estan integrades dins del fitxer en qüestió), recordeu, però, que el propi sistema de fitxers incorpora l'ús d'unes quantes metadades genèriques per qualsevol tipus de fitxer (tipus -si és un arxiu regular, un directori, un enllaç, ...-, mida, permisos, propietari, temps de darrera modificació/accés/canvi/creació, ruta dins de l'arbre de directoris, etc). Aquestes metadades, recordem, són guardades, però, fora del fitxer en sí; són guardades en una estructura interna gestionada pel sistema de fitxers anomenada "inode" (a Windows, MFT) que guarda aquestes metadades i les vincula amb les adreces dels blocs de disc que formen el fitxer en qüestió. La manera més fàcil de conéixer el valor d'aquestes metadades "genèriques" d'un fitxer és mitjançant la comanda *stat*, ja coneguda.

EXERCICIS:

Exiftool (https://exiftool.org) és una comanda de terminal que permet llegir i modificar les metadades incrustades dins d'imatges, vídeos, àudios i documents PDF (la llista completa de formats de fitxers admesos per aquesta eina es troba a la seva pàgina del manual). Aquestes metadades normalment estan incrustades dins dels fitxers en un format anomenat "EXIF" -d'aquí el seu nom- però l'eina Exiftool actualment entén molts altres formats diferents de metadades com poden ser "IPTC", "XMP", etc (la llista completa de formats de metadades es troba a la seva pàgina del manual)

NOTA: Un "wrapper GUI" d'Exiftool és https://github.com/hvdwolf/jExifToolGUI
NOTA: Un programa, també lliure, molt similar en característiques i capacitats a ExifTool és **Exiv2** (https://exiv2.org)
NOTA: Un altre programa lliure que permet obtenir les metadades de fitxers de múltiples formats, que té la possibilitat de treballar en xarxa i que pot mostrar una interfície gràfica (entre altres possibilitats) és **Tika** (https://tika.apache.org)

1.-Instal·la el paquet "libimage-exiftool-perl" (en Ubuntu) o "perl-Image-ExifTool" (en Fedora). Després de llegir l'ajuda bàsica de la comanda *exiftool* mostrada a continuació, utilitza-la per...:

exiftool -s fitxer.png : Mostra les metadades del fitxer indicat. Si no s'indica el paràmetre *-s*, en lloc del nom de les metadades apareixerà una breu descripció, possiblement traduïda. Es pot afegir el paràmetre ***-G1*** per indicar la "categoria" a la que pertany cada metadada ("System", "File", "PNG", "IFD0", "ExifIFD", etc) o ***-G2*** (similar però per categories més genèriques: "Time", "Image", "Camera", etc). També es pot afegir el paràmetre ***-a*** (per mostrar les eventuals metadades amb noms repetits) i ***-u*** (per mostrar els valors també de metadades desconegudes)

NOTA: Cal insistir que no totes les metadades són iguals: algunes són efectivament de tipus EXIF (si n'hi ha), que són les que apareixen a la categoria "G1" anomenada "Exif", però n'hi poden haver de molts altres tipus i formats. I hi ha algunes que són intrínseques al fitxer en sí, com ara la seva mida en bytes, la resolució de la imatge, etc (les quals solen estar incloses dins de les categories "File", "System" o "PNG"/"JFIF"/...),

exiftool -nomMetadada fitxer.png : Mostra (només) el valor actual de la metadada indicada

NOTA: Com a part del nom de la metadada es pot indicar el comodí "*"

Com ja s'ha dit, cada metadada pertany a una categoria "G1"; si a la comanda anterior no s'hi especifica, totes les metadades anomenades igual que puguin pertànyer a categories diferents seran tingudes en compte. Per restringir la sortida de la comanda anterior només a la metadada d'una categoria en concret (que pot ser "G1" o "G2", etc), cal executar llavors la comanda ***exiftool -nomCategoria:nomMetadada fitxer.png***

NOTA: Per conéixer els noms possibles que podria tenir "nomCategoria", a banda dels que ja apareguin fent una exploració del fitxer en sí, es pot executar ***exiftool -listg1*** o ***exiftool -listg2***, etc. Per conéixer totes les metadades pertanyents a una determinada categoria, es pot executar ***exiftool -list -nomCategoria:all*** i per conéixer només les metadades "escrivibles" pertanyents a una determinada categoria, es pot executar ***exiftool -listw -nomCategoria:all***
NOTA: D'altra banda, es pot executar ***exiftool -nomCategoria:all fitxer.png*** per veure totes les metadades de <u>només</u> la categoria indicada

exiftool -nomMetadada=valor fitxer.png : Sobreescriu el valor de la metadada indicada pel valor indicat (o, si no existeix la metadada en qüestió, l'afegeix). El nom de la metadada, en qualsevol cas, no pot ser arbitrari sinó que ha de ser algun dels noms possibles següents: https://exiftool.org/TagNames/index.html També cal tenir en compte, tal com es mostra a la llista anterior, que no totes les metadades són de tipus lectura-escriptura: algunes són de només lectura.

NOTA: Igual que en els casos anteriors, es pot afegir el nom d'una categoria (amb la mateixa sintaxi *-nomCategoria:nomMetadada*) si es vol restringir la metadada en qüestió només a la pertanyent a la categoria indicada

exiftool -all= fitxer.png : Elimina el valor de totes les metadades no imprescindibles del fitxer indicat

NOTA: Igual que en els casos anteriors, es pot afegir el nom d'una categoria (amb la mateixa sintaxi *-nomCategoria:nomMetadada*) si es vol restringir la metadada en qüestió només a la pertanyent a la categoria indicada

---

NOTA: Altres paràmetres interessants d'aquesta eina són (es recomana molt, en tot cas, consultar *man exiftool*):

-TagsFromFile fitxerOriginal (per traspassar les metadades del fitxer indicat a un altre)
-w i ***-htmlDump*** (per escriure les metadades en un fitxer de text o HTML, respectivament)
-t o ***-S*** o ***-s3*** (per modificar el format de la sortida de la comanda -format tabular, "mínim" o només el valor sense indicar el nom de la metadada en qüestió-, respectivament-). També es pot afegir el paràmetre ***-sort*** (per ordenar la sortida alfabèt.)
-r carpeta (per indicar, en lloc d'un fitxer concret, una carpeta el contingut de la qual es processarà recursivament)

D'altra banda, es poden consultar exemples d'ús d'aquesta eina a https://conpilar.es/como-instalar-y-usar-exiftool-en-linux , a https://adamtheautomator.com/exiftool , a https://ninedegreesbelow.com/photography/exiftool-commands.html i també a la seva pàgina oficial, https://exiftool.org/examples.html

a) Comprovar quines són les metadades de la imatge que estàs fent servir actualment com a fons de pantalla. ¿N'hi ha alguna que sigui de tipus EXIF?

b) Copiar aquesta imatge a un altre lloc: ¿la còpia manté les mateixes metadades i els mateixos valors?

c) Trobar una metadada qualsevol de la còpia feta al pas anterior <u>que sigui de tipus lectura-escriptura</u>,,,.

PISTA: Això ho podries veure o bé buscant dins de la categoria "G1" adient dins de la web oficial https://exiftool.org/TagNames/index.html o bé fent servir directament el paràmetre *-listw* de la comanda *exiftool*, explicat al quadre anterior

cII) ... i modificar manualment el valor de la metadada trobada a l'apartat anterior i comprovar que, efectivament, s'hagi canviat.

d) Eliminar la metadada modificada al pas anterior. Tot seguit, afegeix-la de nou amb un nou valor que tu vulguis.

e) Afegir les metadades necessàries per geolocalitzar (falsament) el fitxer-còpia que estàs fent servir en els apartats anteriors (es deixa com a investigació les metadades concretes que hauràs d'emprar). Comprova que, efectivament, s'hagi aconseguit anant a https://tool.geoimgr.com

f) Eliminar totes les metadades del fitxer-còpia que has fet servir en els apartats anteriors, i comprova que, efectivament, s'hagin esborrat

g) Comprovar quines són les metadades d'algun fitxer PDF que tinguis a mà i modifica (o afegeix, si no hi és) les metadades "Author" (persona), "Creator" (software) i "Title" amb el valor que vulguis. Comprova que ho has fet bé i que s'ha generat un fitxer amb extensió ".pdf_original" que representa el PDF original sense modificar (si això no es vol, es pot afegir el paràmetre *-overwrite_original*

> NOTA: A https://exiftool.org/TagNames/PDF.html hi ha la llista completa de metadades PDF que *exiftool* pot manipular. En tot cas, si es coneix l'autor d'un document, es podrien fer atacs de "phishing" personalitzats; si es coneix el software creador es poden buscar vulnerabilitats associades, etc; per tant, no és un tema baladí conèixer les metadades d'un PDF

Més enllà d'eines genèriques com *Exiftool*, també existeixen eines específiques per treballar amb metadades incrustades en formats concrets. A l'exercici següent es mencionen unes quantes:

2.-a) Instal·la el paquet "poppler-utils" per poder executar la comanda *pdfinfo fitxer.pdf* (on "fitxer.pdf" pot ser un fitxer PDF qualsevol). Digues tres metadades de les que se't mostri el seu valor

> NOTA: Eines similars a Poppler que també permeten estudiar/manipular arxius PDF són **Qpdf** (https://github.com/qpdf/qpdf), **Mutool** (https://mupdf.readthedocs.io/en/latest/mutool-info.html) o **PDFtk** (https://gitlab.com/pdftk-java/pdftk), entre altres. En tot cas, a https://github.com/zbetcheckin/PDF_analysis n'hi ha una llista prou extensa. En aquest sentit, a https://gist.github.com/hubgit/6078384 es mostren diverses maneres interessants d'"anonimitzar" (és a dir, eliminar totes les metadades) de documents PDF amb alguna de les eines anteriors.

b) Obre un document ODT qualsevol amb el LibreOffice Writer i vés al menú "File->Properties": allà veuràs (i podràs canviar, o bé una per una manualment o bé mitjançant totes de cop amb el botó "Reset") les metadades del document a la pestanya "General", "Description" i "Custom properties"

> NOTA: També és interessant d'observar el que es veu als quadres que apareixen en clicar sobre les opcions de menú "File->Versions" i "Edit->Changes" (aquesta darrera només tindrà sentit, però, si s'ha activat prèviament l'opció "Record changes" de la pestanya "Security" del quadre "File->Properties")

c) Instal·la el paquet "mediainfo" per poder executar la comanda *mediainfo fitxer.avi* (on "fitxer.avi" pot ser un fitxer multimèdia qualsevol -d'àudio o de vídeo-). Digues 3 metadades de les que se't mostri el seu valor

cII) Vés a https://mediaarea.net/MediaInfoOnline i puja-hi el mateix fitxer multimèdia utilitzat a l'apartat anterior. ¿Quina informació obtens?

> NOTA: Eines similars a MediaInfo que també permeten estudiar/manipular arxius multimèdia són **Ffmpeg** (https://ffmpeg.org), per exemple mitjançant la comanda *ffmpeg -i fitxer.avi -map_metadata -1 -c:v copy -c:a copy* o bé la comanda *ffmpeg -i fitxer.avi -v info -f null -* . En el cas concret de les fotografies, també es poden usar la comanda *gm identify* (del paquet **GraphicsMagick**, http://www.graphicsmagick.org) o també *pngcheck* (del paquet homònim)

3.-a) Vés a https://hintfo.com i vés a la pestanya "Metadata". ¿Quina diferència hi ha entre el que allà s'anomenen metadades "implícites", "explícites" i "externes"?

b) Puja una imatge mitjançant el botó que t'ofereix aquesta web i compara les dades obtingudes per Exiftool, Exiv2 i MediaInfo. ¿Quina eina ofereix en aquest cas una informació més completa?

Un programa lliure que permet eliminar les metadades de fitxers de múltiples formats és **Mat2** (https://0xacab.org/jvoisin/mat2). A l'exercici següent es demana provar la interfície gràfica (https://gitlab.com/rmnvgr/metadata-cleaner)

> NOTA: A https://miloserdov.org/?p=3130 hi ha un tutorial de Mat2 prou interessant aplicat a documents LibreOffice
> NOTA: Una altra eina similar a Mat2 i també a Exiftool és https://hachoir.readthedocs.io/en/latest

4.-a) Executa la comanda *flatpak install --user flathub fr.romainvigier.MetadataCleaner* per instal·lar el programa "Metadata Cleaner" i executa'l des del llençador de l'escriptori o bé directament amb la comanda *flatpak run fr.romainvigier.MetadataCleaner* Utilitza'l per eliminar totes les metadades d'algun document ODT, PDF o imatge PNG/JPG (el que vulguis)

b) Elimina totes les metadades d'algun document ODT, PDF o imatge PNG/JPG (el que vulguis) fent servir la versió web de mat2: https://metadata.systemli.org (el codi de la qual és lliure i es pot trobar a https://0xacab.org/jvoisin/mat2-web)

5.-a) ¿Què explica aquest article https://exposingtheinvisible.org/en/guides/image-digging?

> NOTA: Per saber més tècniques d'anàlisi forense de fotografies, on intervenen metadades i altres elements, recomano la lectura de l'article https://blog.elhacker.net/2016/12/analisis-forense-en-imagenes-digitales-metadatos-prnu.html

b) ¿Què explica aquest article https://miloserdov.org/?p=5621?

D'entre les metadades classificades com a "externes" a l'exercici 3, podem destacar els diferents tipus de dates (en anglès, "timestamps" o "marques de temps") associades a tots els fitxers d'un sistema.

> NOTA: Aquestes metadades es guarden físicament, en el cas de Linux, a l'anomenada "taula d'inodes" (on cada "inode" representa l'entrada corresponent a un fitxer o directori concret), o bé, en el cas de Windows, a l'anomenada MFT. Ambdues estructures es troben al llarg del sistema de fitxers en qüestió i són imprescindible per localitzar els diferents sectors on cada fitxer (definit per les seves metadades) està ubicat físicament al llarg del sistema de fitxers.

Hora de la darrera modificació	Per als directoris, aquesta és l'última vegada que es va afegir, canviar el nom o eliminar una entrada. Per a altres tipus de fitxers, és l'última vegada que es va escriure el fitxer.
Hora del darrer accés	Per als directoris, aquesta és l'última vegada que es va llistar el seu contingut o s'hi va entrar. Per a altres tipus de fitxers, és l'última vegada que es va llegir el fitxer.
Hora del darrer canvi	Última vegada que es va modificar alguna metadada inclosa a l'inode associat al fitxer o directori en qüestió. Exemples d'aquestes metadades són les que apareixen a la sortida de la comanda *ls -l*: propietari i grup, permisos, nombre d'enllaços durs, qualsevol de les hores MAC ("modify-access-change-creation"), la mida,...
Hora de creació	Alguns sistemes de fitxers registren l'hora en què es va crear el fitxer, però no tots

Per llegir les metadades corresponents a un fitxer en un sistema Linux es pot utilitzar la comanda **stat fitxer**, la qual té, a més, un conjunt de paràmetres interessants que s'estudiaran en els propers exercicis. També és útil la comanda **ls -li fitxer** (el paràmetre *-i* afegeix a la sortida una columna extra amb el nº d'inode corresponent a cada entrada mostrada). Una altra comanda interessant és **df -i** , la qual compta quants inodes hi ha en total i lliures a cada partició reconeguda.

> NOTA: Per defecte, la comanda *ls -l* mostra la data de darrera modificació però aquesta es pot canviar per la de darrer accés (si s'indica el paràmetre **-u**), la de darrer canvi (si s'indica el paràmetre **-c**) o la de creació (si s'indica el paràmetre **--time=creation**). En tot cas, per ordenar, sigui quina sigui la data mostrada, es pot afegir el paràmetre **-t** (el més nou primer) o bé els paràmetres **-tr** (el més antic primer)
> NOTA: Cal tenir en compte que segons com s'hagués muntat el sistema de fitxer original del dispositiu a estudiar mentre es feia servir, alguna de les dates anteriors pot no estar disponible. Concretament, l'opció **noatime** desactiva l'actualització de la data de darrer accés (per tal d'evitar escriptures en disc innecessàries quan només s'està volent fer operacions de lectura sobre el sistema de fitxers). Si aquesta funcionalitat es vol no pel sistema de fitxers complet sinó per fitxers concrets, es pot aconseguir alternativament amb la comanda **chattr +A fitxer**. També està l'opció **nodiratime**, que només desactiva l'actualització de la data de darrer accés pels directoris. D'altra banda, també està l'opció de muntatge **relatime**, la qual és un "compromís": només modifica la data de darrer accés si aquesta és més antiga que la de darrera modificació o la de darrer canvi (o si ha passat més de 24h des de la darrera data d'accés registrada)

6.-a) Fes servir la comanda *stat* per esbrinar la data de darrera modificació, de darrer accés, de darrer canvi de metadades i de creació dels fitxers "/etc/passwd" i "/etc/fstab", a més de la seva mida, permisos, propietari, nombre d'enllaços durs i nombre d'inode respectius.

b) Compara la sortida de les comandes *stat /bin* i *stat -L /bin* ¿A què correspon cadascuna?

c) ¿Quina és la dada concreta que mostra la comanda *stat* (sobre un fitxer o carpeta qualsevol) si li indiquem el paràmetre *-c* acompanyat del valor *%F* (així , *stat -c %F unfitxer*). ¿I acompanyat del valor *%n*? ¿I del valor *%a?* ¿I del valor *%A*? ¿I de *%s*? ¿I de *%i*? ¿I de *%w %x %y %z*?

d) Compara les dades obtingudes a l'apartat a) amb les mostrades per les comandes *ls -il /etc/passwd /etc/fstab* , *ls -ilu /etc/passwd /etc/fstab* i *ls -ilc /etc/passwd /etc/fstab* ¿Coincideixen?

e) Executa *cat hola > a.txt* i tot seguit observa les dates d'última modificació, canvi, accés i creació d'aquest fitxer (amb *ls -l a.txt; ls -lc a.txt; ls -lu a.txt i ls -l --time=creation a.txt* respectivament, o bé amb *stat a.txt* d'un sol cop). Tot seguit, executa la comanda *touch* ***-m*** *-d "2023-10-25 23:54:21.456" a.txt* i observa què li ha passat a la data d'última modificació del fitxer en qüestió. Finalment, executa la comanda *touch* ***-a*** *-d "2023-10-25 23:54:21.456" a.txt* i observa què li ha passat a la seva data d'últim accés.

> NOTA: Les dates de creació i d'últim canvi no es poden editar amb la comanda *touch* perquè només hi té accés el sistema operatiu a baix nivell. Una alternativa, si volem comprometre també aquesta informació temporal per protegir-nos, en cas de ser intrusos, d'un eventual anàlisi forense del timeline (aquesta tècnica d'evasió, coneguda com "timestomp" és molt popular, tal com es pot veure aquí: https://attack.mitre.org/techniques/T1070/006) és alterar manualment la data general del sistema (segurament caldrà desactivar la sincronització NTP abans), implantar llavors els fitxers en qüestió dins de la ubicació desitjada i finalment tornar el sistema a la data correcta.

f) ¿Quina diferència hi ha entre la sortida de la comanda *df -h* i *df -i*?

g) Esbrina el número d'inode associat al teu arxiu ".bashrc" (suposarem que és 175643, per exemple) i tot seguit executa: *find / -inum 175643 2> /dev/null* ¿Per a què serveix aquesta comanda concreta i quin resultat obtens?

> NOTA: En el cas d'estar en un sistema amb múltiples particions muntades (és a dir, múltiples sistemes de fitxers diferents), podria passar que la comanda anterior trobés més d'un element amb el mateix nombre d'inode (ja que com que aquest és particular per cada sistema de fitxer, es podria donar el cas que hi hagués aquesta coincidència)