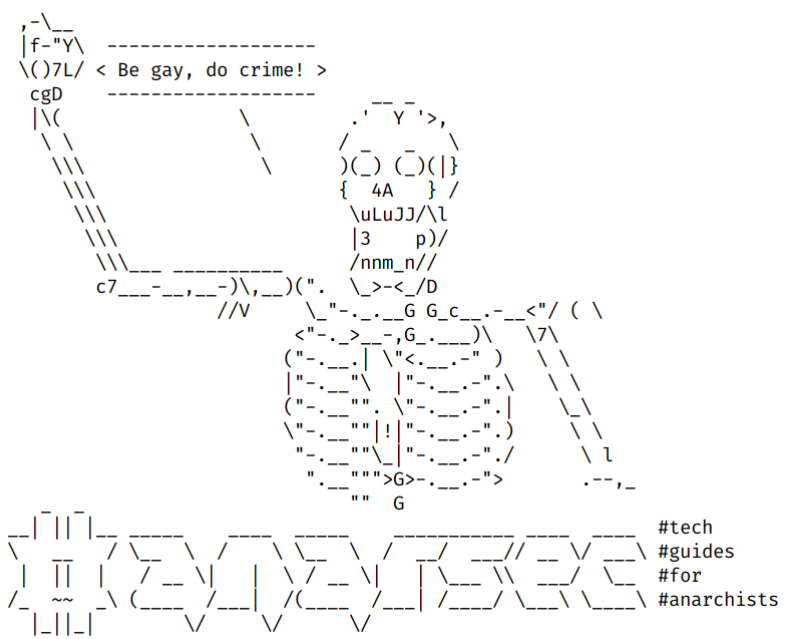
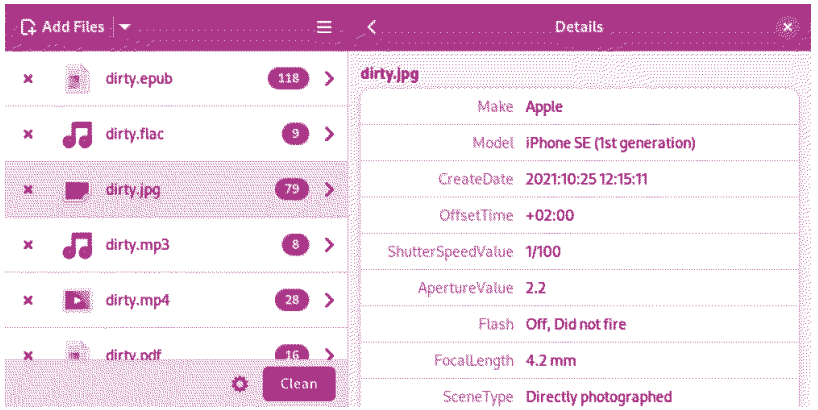
Remove Identifying Metadata From Files

Metadata is data about data or information about information. In the context of files, this can mean information that is automatically embedded in the file, and this information can be used to deanonymize you. For example, an image file will often have metadata about when it was taken, where it was taken, what camera it was taken with, etc. A PDF file may have information about what program created it, what computer, etc. This can be used by investigators to link a photo to the camera on which it was taken, a video to the computer on which it was edited, and so on. Before you put a sensitive file on the Internet, remove the metadata.





Series: Defensive

This version of the zine was last edited on 2024-04-20. Visit anarsec.guide to see whether it has been updated since.

The dagger symbol † on a word means that there is a glossary entry for it. Ai ferri corti.

AnarSec is a resource designed to help anarchists navigate the hostile terrain of technology — defensive guides for digital security and anonymity, as well as offensive guides for hacking. All guides are available in booklet format for printing and will be kept up to date.

Defensive

Tails

- Tails for Anarchists
- Tails Best Practices

Qubes OS

- Qubes OS for Anarchists

Phones

- Kill the Cop in Your Pocket
- GrapheneOS for Anarchists

General

- Linux Essentials
- Remove Identifying Metadata From Files
- Encrypted Messaging for Anarchists
- Make Your Electronics Tamper-Evident

Offensive

Coming soon

Contents

Metadata Anonymization Toolkit 4
Using the Metadata Cleaner 4
Photo and Video Forensics 5
Printer Forensics 6
Further Reading 6
Appendix: Recommendations 7
    Your Phone 7
    Your Computer 8
    Encrypted Messaging 9
    Storing Electronic Devices 9
Appendix: Glossary 9
    Command Line Interface (CLI) 9
    Metadata 9
    Operating system (OS) 10
    Tor network 10

Metadata† is *data about data* or *information about information*. In the context of files, this can mean information that is automatically embedded in the file, and this information can be used to deanonymize you. For example, an image file will often have metadata about when it was taken, where it was taken, what camera it was taken with, etc. A PDF file may have information about what program created it, what computer, etc. This can be used by investigators to link a photo to the camera on which it was taken, a video to the computer on which it was edited, and so on. Before you put a sensitive file on the Internet, remove the metadata.

Metadata Anonymization Toolkit

Fortunately, there is a tool that comprehensively cleans metadata, and it is available as both a command line interface† and a graphical user interface. The command line version is called `mat2` and is open-source[1], and the graphical version is called Metadata Cleaner and is also open-source[2]. Both programs are included in Tails[3] and Qubes-Whonix[4] by default.

Using the Metadata Cleaner

If you are not comfortable with the command line, we recommend using Metadata Cleaner — it uses `mat2` under the hood, so it has all the same functionality. Metadata Cleaner is better than Exiftool and other metadata removal software — see the comparison docs[5].

[1] 0xacab.org/jvoisin/mat2
[2] gitlab.com/rmnvgr/metadata-cleaner/
[3] anarsec.guide/tags/tails/
[4] anarsec.guide/posts/qubes/#whonix-and-tor
[5] 0xacab.org/jvoisin/mat2/-/blob/master/doc/comparison_to_others.md

Metadata Cleaner shows the metadata it detects, but "it doesn't mean that a file is clean from any metadata if mat2 doesn't show any. There is no reliable way to detect every single possible metadata for complex file formats." This means that you should clean the file even if no metadata is displayed.

To use the Metadata Cleaner, first add a file. When you click it, the current metadata is displayed. Select the file, then select **Clean**. You can verify that the metadata has been removed by re-adding the cleaned file and viewing its metadata.

When you clean a PDF file, it is converted to images, so the quality is downgraded and you cannot select the text in it. If you want to retain this ability, there is a *lightweight* cleaning mode that cleans only the superficial metadata of your file, but not the metadata of "embedded resources" (such as images in the PDF). If you are creating a PDF, use Metadata Cleaner on any images before importing them into the layout software, and use layout software on Tails or Qubes-Whonix such as Scribus that are generic for those operating systems. You can enable "lightweight cleaning" in the Metadata Cleaner settings.

Note the limitations of Metadata Cleaner: "mat2 only removes metadata from your files, it does not anonymise their content, nor can it handle watermarking, steganography, or any too custom metadata field/system. If you really want to be anonymous, use file formats that do not contain any metadata, or better: use plain-text."

Photo and Video Forensics

While it is possible to remove all metadata from an image or video, forensic examination may still reveal what device was used to capture it. As the Whonix docs[6] note:

[6] whonix.org/wiki/Surfing_Posting_Blogging#Photographs

> Every camera's sensor has a unique noise signature because of subtle hardware differences. The sensor noise is detectable in the pixels of every image and video shot with the camera and could be fingerprinted. In the same way ballistics forensics can trace a bullet to the barrel it came from, the same can be accomplished with adversarial digital forensics for all images and videos. Note this effect is different from file metadata.

Multiple photos or videos from the same camera can be tied together in this way, and if the camera is recovered, it can be confirmed where the files came from. Cheap cameras can be purchased from a pawn shop and used only once for pictures or videos that require high security.

Printer Forensics

All modern printers leave invisible watermarks to encode information such as the serial number of the printer and when it was printed. When printed material is scanned, these marks are present in the file. To learn more, see Revealing Traces in Printouts and Scans[7] and the Whonix documentation on printing and scanning[8].

Further Reading

- Anonymous File Sharing[9] from the Whonix documentation.
- Redacting Documents/Pictures/Videos/Audio safely[10] for a table of recommended software for creating different types of files.
- Behind the Data: Investigating metadata[11] for how metadata can be used to identify and reveal personal information.

[7] dys2p.com/en/2022-09-print-scan-traces.html
[8] whonix.org/wiki/Printing_and_Scanning
[9] whonix.org/wiki/Surfing_Posting_Blogging#Anonymous_File_Sharing

For more information, see Tails for Anarchists[32] and Privacy Guides[33]. To understand the limitations of Tor, see the Whonix documentation[34].

[32] anarsec.guide/posts/tails/#tor
[33] privacyguides.org/en/advanced/tor-overview/
[34] whonix.org/wiki/Warning

target is using), government agencies (to collect information about people to create social graphs), and other parties (to target location-based advertising). Whenever you use a computer, you are likely leaving metadata behind.

For more information, see Remove Identifying Metadata From Files[29] and Defend Dissent: Metadata[30].

Operating system (OS)

The system software that runs your device before any other software. Some common examples include Windows, macOS, Linux, Android, and iOS. Linux and some versions of Android are the only open-source options on this list.

Tor network

Tor[31] (short for The Onion Router) is an open and distributed network that helps defend against traffic analysis. Tor protects you by routing your communications through a network of relays run by volunteers around the world: it prevents someone monitoring your Internet connection from learning what sites you visit, and it prevents the operators of the sites you visit from learning your physical location.

Every website visited through the Tor network passes through 3 relays. Relays are servers hosted by different people and organizations around the world. No single relay ever knows both where the encrypted connection is coming from and where it is going. An excerpt from a leaked top-secret NSA assessment calls Tor "the King of high secure, low latency Internet anonymity" with "no contenders for the throne in waiting". The Tor network can be accessed through the Tor Browser on any operating system. The Tails† operating system forces every program to use the Tor network when accessing the Internet.

[29] anarsec.guide/posts/metadata

[30] open.oregonstate.education/defenddissent/chapter/metadata/

[31] torproject.org/

Appendix: Recommendations

As anarchists, we must defend ourselves against police and intelligence agencies that conduct targeted digital surveillance[12] for the purposes of incrimination[13] and network mapping[14]. Our goal is to obscure the State's visibility into our lives and projects. Our recommendations are intended for all anarchists, and they are accompanied by guides to put the advice into practice.

We agree with the conclusion of an overview of targeted surveillance measures in France[15]: "So let's be clear about our responsibilities: if we knowingly bring a networked device equipped with a microphone and/or a camera (cell phone, baby monitor, computer, car GPS, networked watch, etc.) close to a conversation in which "private or confidential words are spoken" and must remain so, even if it's switched off, we become a potential state informer..."

You may also be interested in the Threat Library's "Digital Best Practices"[16].

Your Phone

Operating system†: GrapheneOS is the only reasonably secure choice for cell phones. See GrapheneOS for Anarchists[17]. If you decide to have a phone, treat it like an

[10] anonymousplanet.org/guide.html#redacting-documentspicturesvideosaudio-safely

[11] exposingtheinvisible.org/en/guides/behind-the-data-metadata-investigations/

[12] notrace.how/threat-library/techniques/targeted-digital-surveillance.html

[13] notrace.how/threat-library/tactics/incrimination.html

[14] notrace.how/threat-library/techniques/network-mapping.html

[15] actforfree.noblogs.org/post/2023/07/24/number-of-the-day-89502-preventive-surveillance-measures-france/

[16] notrace.how/threat-library/mitigations/digital-best-practices.html

"encrypted landline" and leave it at home when you are out of the house. See Kill the Cop in Your Pocket[18].

Your Computer

Operating system†: Tails is unparalleled for sensitive computer use (writing and sending communiques, moderating a sketchy website, researching for actions, reading articles that may be criminalized, etc.). Tails runs from a USB drive and is designed with the anti-forensic property of leaving no trace of your activity on your computer, as well as forcing all Internet connections through the Tor network†. See Tails for Anarchists[19] and Tails Best Practices[20].

Operating system†: Qubes OS has better security than Tails for many use cases, but has a steeper learning curve and no anti-forensic features. However, it is accessible enough for journalists and other non-technical users. Basic knowledge of using Linux is required — see Linux Essentials[21]. Qubes OS can even run Windows programs such as Adobe InDesign, but much more securely than a standard Windows computer. See Qubes OS for Anarchists[22].

See When to Use Tails vs. Qubes OS[23]. We do not offer "harm reduction" advice for Windows or macOS computers, as this is already widespread and gives a false sense of privacy and security.

[17] anarsec.guide/posts/grapheneos/
[18] anarsec.guide/posts/nophones/
[19] anarsec.guide/posts/tails/
[20] anarsec.guide/posts/tails-best/
[21] anarsec.guide/posts/linux
[22] anarsec.guide/posts/qubes/
[23] anarsec.guide/posts/qubes/#when-to-use-tails-vs-qubes-os

Encrypted Messaging

See Encrypted Messaging for Anarchists[24]

Storing Electronic Devices

See Make Your Electronics Tamper-Evident[25].

Appendix: Glossary

Command Line Interface (CLI)

The "command line" is an all-text alternative to the graphical "point and click" tool that most of us are more familiar with; the Command Line Interface (CLI) allows us to do some things that a Graphical User Interface (GUI) does not. Often, either a GUI or a CLI would work, and which you use is a matter of preference. For example, in Tails†, you can verify the checksum[26] of a file using either a GUI (the GtkHash program) or a CLI command (`sha256sum`).

For more information, see Linux Essentials[27]. The Tech Learning Collective's "Foundations: Linux Journey" course on the command line[28] is our recommended introduction to using the CLI/terminal.

Metadata

Metadata is data that provides information about other data. For example, a JPG file contains the actual image (data) but it may also contain metadata such as the date the file was created, the type of camera, GPS coordinates, and so on. Metadata can be valuable to attackers (to find appropriate exploits for outdated software the

[24] anarsec.guide/posts/e2ee/
[25] anarsec.guide/posts/tamper/
[26] anarsec.guide/glossary
[27] anarsec.guide/posts/linux/#the-command-line-interface
[28] techlearningcollective.com/foundations/linux-journey/the-shell