Source: Bing Chat AI

INTRODUCTION TO OSINT

DEPARTMENT OF COMPUTER SCIENCE. UNIVERSITY OF OVIEDO

Open-Source Research Techniques | 2023 – 2024 (V1.5 "*A-21 'Poseidon' Lite* ")

José Manuel Redondo López

Content

What is this and what is it for? **6**

Introduction ***6***

The doxing crime ***7***

What kind of information is collected? ***8***

To "engage" or "to not engage" that is the question ***10***

Passive techniques 10

Active techniques 10

The "Intelligence Cycle" 11

"Personal" Anonymity: Creating a False Online Identity ***12***

Use of temporary emails to register on services 12

Creation of fake identities 28

What if I need more than just personal data to create a fake identity? 36

Machine anonymity ***38***

Remote Browsing 38

Program sandboxing 38

VPNs 39

Connections via ToR 40

Searches from another location 44

Preparing the OSINT study. *What kind of information should we look for?* **45**

"Manual" OSINT: The art of searching and interpreting data **51**

OSINT on employees or individuals ***51***

Things about persons in general: Names / Images / Videos / Photos... 52

Email Addresses / Domain Names 62

Social Media 65

People search networks 76

Dating Apps 77

Phone numbers 78

Public Records 82

Telephones and credit cards (or other documents of which pictures appear on the Internet) 86

Use of various means of transport 87

Geolocation 88

Forums / Blogs / IRC... 91

Announcements 92

OSINT on machines and code ***94***

Mobile terminal emulation 94
IP Addresses / Domains 95
Offline analysis of files on the network 96
What happens if you leave your devices unlocked 98

"Automatic" OSINT. OSINT Ex Machina 100

OSINT on people 100

OSINT on machines: Mass metadata with FOCA 102

Introduction 102
Installation 103
Using FOCA 103

Data Mining for OSINT: Maltego 109

Installation 109
First boot 111
Creating a project 115
Automated information search 117
Detail of a *Maltego* entity 124
Creating graphs manually 126
Searching on social networks 130
Other Maltego entities 131
The Transformation Center 132
Export the information 137
I want to know more! 138

"Automatic" OSINT: Search Engines and Threat Intelligence 141

Introduction 141

OSINT APIs: OSINT in our programs! 143

Interactive OSINT tool trees classified by performed task 145

Malfrat's OSINT (https://map.malfrats.industries/) 145
OSINT Framework (https://osintframework.com/) 145

Search engines for machines and networks 147

Binary Edge (https://www.binaryedge.io/) 148
Censys (https://search.censys.io/) 148
DNS: Specific tools for servers and domains 150
FOFA (https://en.fofa.info/) 151
FullHunt (https://fullhunt.io/) 152
GreyNoise (https://viz.greynoise.io/) 153
IVRE (https://ivre.rocks/) 154
Netlas (https://app.netlas.io/responses/) 155
OnHype (https://www.onyphe.io/) 156

PulseDive (https://pulsedive.com/) 157
Shodan (https://www.shodan.io/) 158
Wigle (https://wigle.net/) 160
ZoomEye (https://www.zoomeye.org/) 161
Search engines for websites or networks 162
Advance Online Utilities (https://centralops.net/co/) 162
Cisco Reputation Check (https://www.talosintelligence.com/) 162
Cloudflare Radar URL Scanner (https://radar.cloudflare.com/scan) 163
CyberGordon (https://cybergordon.com/) 163
IBM Reputation Check (https://exchange.xforce.ibmcloud.com/) 164
IP Reputation Check (https://www.abuseipdb.com/) 164
PublicWWW (https://publicwww.com/) 165
URL Category Finder (https://sitereview.bluecoat.com/#/) 165
URLScan (https://urlscan.io) 166
Website Profiler Tool (https://builtwith.com/) 167
Vulnerability search engines 168
CyberChef (https://gchq.github.io/CyberChef/) 168
Packet Storm Security (https://packetstormsecurity.com/) 168
Vulners (https://vulners.com/) 169
Search engines for general data breaches 171
CRT.sh (https://ui.ctsearch.entrust.com/ui/ctsearchui) 171
FilePursuit (https://filepursuit.com/) 171
Gray Hat Warfare (https://buckets.grayhatwarfare.com/) 172
IntelligenceX (https://intelx.io/tools) 173
LeakIX (https://leakix.net/) 174
TinEye Reverse Image search (https://tineye.com) 175
Specific search engines for leaks related to user accounts 177
Hunter.io (https://hunter.io/) 177
LeakCheck (https://leakcheck.io/) and Have I been pwnd? (https://haveibeenpwned.com/) 177
Search engines for source or executable code 179
Grep.App (https://grep.app/) 179
Joe SandBox (https://www.joesandbox.com/) 179
SearchCode (https://searchcode.com/) 180
Virustotal (https://www.virustotal.com/gui/home/upload) and VT Intelligence (https://www.virustotal.com/gui/intelligence-overview) 182
When the end justifies the means 184
Phishing and Spear Phishing to obtain information 184
SET 185
Gophish 196

Evilgnx2 197

Dark Web Investigations ***197***

😇 Addendum: A friend wants to know more about these topics (okay, the friend is me): Sites where you can learn more about OSINT **200**

Guides & Articles ***200***

Documentation repositories, guides, and tools ***200***

To defeat an enemy, you must know them

Not only their battle tactics

But the history, philosophy, art… […]

So, we can always be one step ahead of them

Admiral Thrawn

What is this and what is it for?

Introduction

OSINT stands for ***Open-Source Intelligence*** (intelligence on open information sources, that is, extracting information from "things" that one finds on the Internet) and is **one of the key aspects** to understand the cybersecurity that governs the Internet today. Some students called it last year "being a cibermaruja", and I think it is a definition that captures the basic idea 😊.

Doing research with this type of open information implies **asking many questions about individuals or entities to be studied (our objectives) and giving a more or less precise answer to them**. For this, a series of activities must be done that, in the eye of a person who believes that security has a purely technical component, can be surprising or "not serious". **But nothing could be further from the truth**.

In an increasingly interconnected world, crimes based on the misuse of exposed information (usually improperly) will not go down, quite the opposite. These crimes will happen to people, to companies, and on the software products they execute. No one is spared, and the only "vaccine" is to be as educated and alert as possible, although nothing is foolproof. Working with this kind of information **is the favorite means to gather information that triggers very effective phishing and/or spear phishing attacks**, which have brought down giants. An example is the February 2023 Reddit breach. And many, so many others that have passed and will happen.

Why? Because we have concentrated our efforts on defending machines and infrastructures, which is very good **and very necessary**. But perhaps we have not put the same effort into defending people. The people operate those machines and infrastructure, have an extremely variable technical knowledge profile, **and they are the weakest link**. Human error is responsible for 82% of data breaches (https://www.grcelearning.com/blog/human-error-is-responsible-for-85-of-data-breaches). Therefore, in this part of the course we are going to dedicate a section to people, since in the others we are focusing on machines (and code!). We believe that this way you will be much better trained for the future.

And this is a security discipline that is very useful and attractive, with many ramifications that touch on a lot of technical and also social aspects, usable from the "defensive" and the "offensive" side. This makes it a very attractive aspect of security for many people, and it is my intention to use this document as a "core" **to develop a complete course on this subject (the A-21 "Poseidon", but not Lite 😊)**, with a more extensive organization and with more concepts, organized to get a sense of this broader and more transversal discipline. But for now, enjoy this "crash course" on OSINT! 😊

The doxing crime

OSINT-based research should not be confused with ***doxing*** (also called *'doxxing'*), an abbreviation for "Drop Documents". This is a **type of cyber-attack** (i.e., a **crime**) in which the attacker **steals** the victim's personal information and leaks it to the internet.

In a sense both operations are related. In both cases we try to obtain as much information as possible about an objective, but in the OSINT case it is done with open sources, or in other words, information whose access is free. This implies that **it is done legally** unless we later use that information to extort or cause some kind of harm to the victim. In the case of doxing, it is done from restricted or non-public sources to which you have access, which makes it **an illegal activity**. *Reason?* Accessing a person's confidential information **is a crime**, unless you have a legal mandate, or the law protects you to do so because it is a very specific case. **Disclosing confidential information then is always a crime.**

In this course we do NOT want criminals, but people who know how they act in order to defend themselves against them. It is very different, even if some people do not think it is. Without knowing the weapons available to the enemy, we hardly can a) establish effective countermeasures and b) detect points vulnerable to the enemy's weapons. Within our possibilities, we have not left anything in this regard in this document. What you see is crudely what there is, and I assure you that **there are much worse things** that I have not included simply because of lack of time or because they are refinements of things that you will see here. This will help you realize how bad we are in the security aspect, and act accordingly in your environment or future work.

Usually, the goal of doxing is to **harass or harm** the victim in some way. The variants depend on the context and the victim, and it is not possible to list them here. But you can get an idea as you read the examples that I give you here and let your imagination fly.

Unfortunately, **doxing is more common than you think**. All kinds of people are potential victims, no matter if you are famous or not. There are "bad" people out there who may want to hurt you or your relatives for some reason (envy, supposed moral superiority, not considering you "worthy", for belonging to a certain group (age, company, ideology, ...), there are thousands of "reasons"...). **This part of computer science links directly to psychiatry or psychology topics**. Or there is no reason, but just for "fun" (this is directly a case study for psychiatry).

To avoid doxing, it is very important **to be very careful about the type of information you put on the Internet**. Sometimes there can be a "controlled" doxing scenario, where a person doing an offensive security audit has the authorization to "steal" certain kind of information that can then be used to continue with an attack chain such as the one commented on by MITRE ATT&CK (**Topic 7**), **although never exposing it outside, use it to harm others or publish it**. We are professionals, **NOT criminals**. Here it is about knowing what could be stolen pretending to be an adversary, not being a "method actor" (*surely you understand what I mean, right?*).

What kind of information is collected?

In any case, whether from public or non-public sources, this means **knowing things about the target** that can then be used for unfolding **attacks** (phishing that involves impersonation or convincing the target of something, finding out credentials, impersonating someone ...) or obtaining **any kind of tactical or strategic advantage** over that objective. This includes aspects such as the following, although we will see more later:

- Real name / last name
- Physical characteristics (weight, height...), hobbies, hobbies, personal tastes...
- Email addresses
- Phone numbers (with which we will have access to the victim via instant messaging)
- Social media profiles
- Address or location
- Jobs (current and past)
- Bank account information, credit cards...
- Opinions on certain issues, ideologies, religious faith...

The "**commandments**" of the intelligence gathering from open sources or OSINT are four:

1. **Open-Source**
2. **Public**
3. **Free**
4. **Legal**

I mean:

- It must be **objective** intelligence, free of bias. For example, if it is the opinion of someone in a social network of something, it must be contrasted with one or several reliable sources to avoid falling into hoaxes or **fake news**.
- Data **must be available from public sources**, such as those available on the Internet, although the term is not limited to the Internet but includes all publicly available sources (physical or digital) where the auditor can obtain the information. It must be available **free of charge**, except in very specific justified cases. It does not matter if the information is inside newspapers, blogs, web pages, tweets, social networks, images, podcasts, videos... as long as it is public, free, and legal.

With the right information we can gain a great advantage in our reconnaissance operations (phase ***TA0043 Reconnaissance*** of the *MITRE ATT&CK*, **Topics 5 and 7**) and accelerate any investigation of the company / people. Although OSINT is usually associated with cyber warfare, cyber-attacks, cybersecurity, etc. (and it is certainly part of it), and therefore seem like really complex actions, the operations to search among this type of information are usually not complicated once the techniques, tools, and what to look for from each found information are known. As an example of what to do, we have these operations:

- Ask many question types to any **search engine**. Yes, this includes modern AI-based chat engines such as ***ChatGPT*** or ***Bing Search***. You can ask for data of a person and surely be able to give you a truthful and coherent answer in many cases.
- Investigate **public forums, blogs, or social networks** (identifying them previously) about the behavior of a certain person in them, to see what things he tells, reveals, "escapes" or admits to knowing believing that he does it anonymously. People often reveal many things in these situations.

- Investigate **public documents** about things that have happened to the objective (scholarship awards, prizes, projects, fines, non-payments ...).
- Investigate if the objectives have any **possession or property** in their name (cars, houses for sale ...)
- **Everyday information** that we believe may serve our purposes.
    - Full names of **employees**, **job roles**, and the **software** they use.
    - **Old data** cached by *Google*/*Wayback Machine* often reveals interesting information.
    - **Phone numbers, email addresses, social networks,** or *Google* results.
    - **Photos and videos** on common social photo sharing sites
    - *Google Maps* and the like to retrieve images of users' **geographical location**.

Data sources will be plenty (look at this image!), now we have to know how to search for them and what to look for

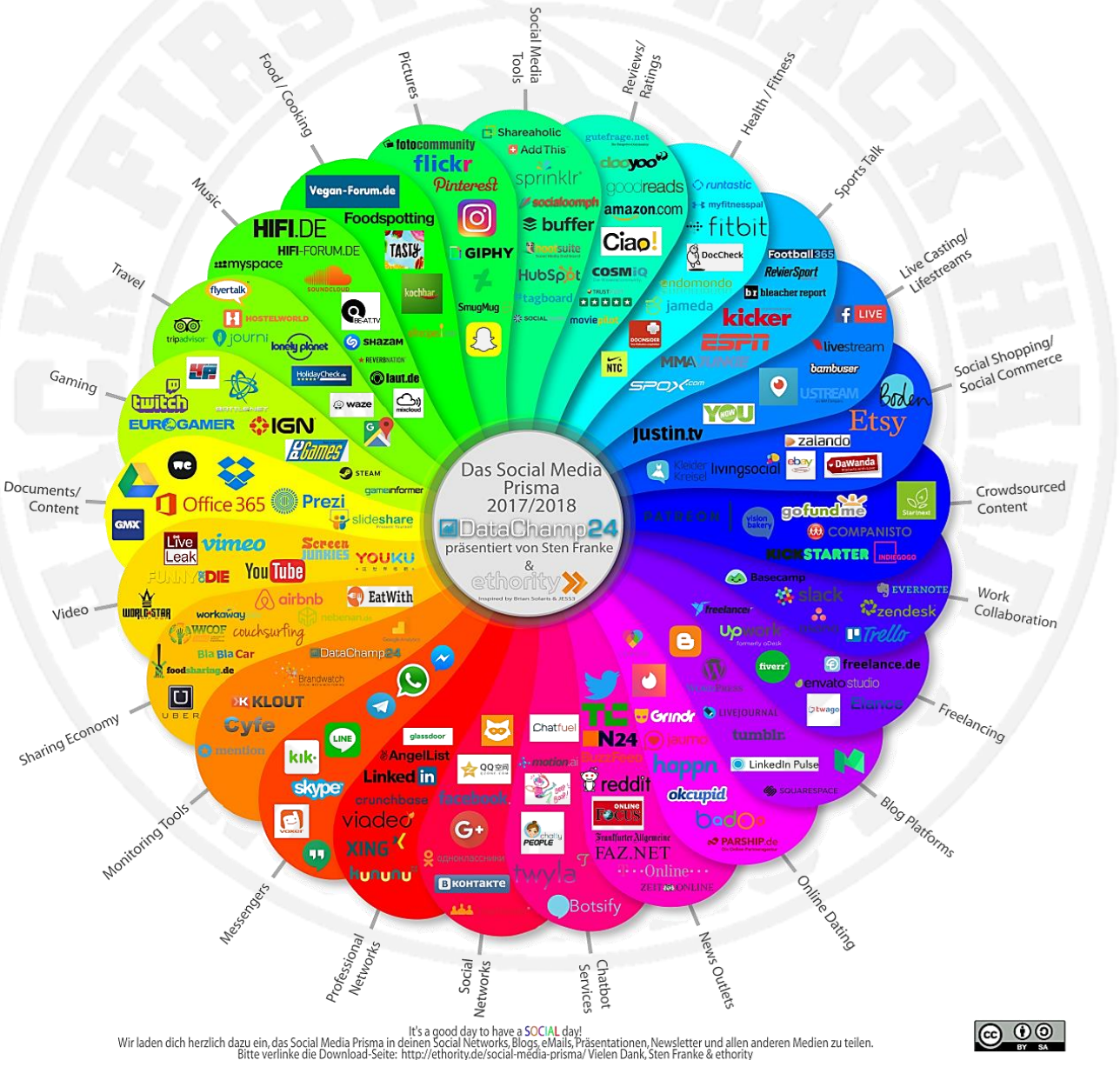


Figure 1. Information sources by type of document or searched data

To "engage" or "to not engage" that is the question

To make OSINT studies there are **two types of techniques**, depending on the degree of interaction made with the objective.

Passive techniques

The main characteristic of these techniques is that **the investigated objective can NEVER know that he is being investigated**, that is, **you SHOULD NEVER contact him / her (or any member of the organization / company)**, under any circumstances. **Not even under a false identity**. This is done when you should not arouse suspicion of any kind in anyone investigated, and thus the "trace" of your investigations will be "submerged" in huge logs of many foreign systems that surely nobody will look at, because in fact you have not done anything illegal or what you can be accused of. **The information was there publicly, and you have only found it**. If there is no criminal use of the information and you give no reason for someone to track your steps, everything you do will be lost, like tears in the rain 😉.

In addition to minimizing the risk of detection, this type of OSINT mostly **uses third-party services to do its work** (for example, search engine results such as *Google* or *Shodan* such as those seen in **Seminar 1** and **Laboratory 2**). In other words, the contact with the objective is not made by you, **but by others for you**, who processes the information, and gives it to you classified and ordered in some way, so that you consume it as you like. Since that third party is normally engaged in doing that, it is rare (though possible) for the target to block it. It also depends on the third party, of course. Blocking research agents ("web spiders") from *Shodan* machines is more common than blocking *Google*, *Bing* or similar ones, unless one decides not to be indexed and belong to the ***Deep Web*** (https://en.wikipedia.org/wiki/Deep_web).

However, this ability to go unnoticed has a price: **greater number of false positives and "noise" in the results** (results that seem to be correct, but when looking at them they turn out to be from another entity or person with a similar name or characteristics).

Active techniques

These types of techniques involve **direct contact with the target**. The most typical example is a port scan with `nmap`, such as those described in **labs 1, 2, and 8-9**. Unless you take **measures to preserve anonymity** at the level of the person or machine doing the operation (which we will see in the next section**), they will know that it is you**. In any case, they will detect you, and therefore, **they can block you** if they consider that what you do is anomalous or offensive. It is all about a race between target detection capabilities and your ability to "hide." In exchange for losing anonymity and increasing the probability of blocking, you get **greater reliability of results** in general and usually **more details of the target**.

Therefore, due to the risk of discovering the auditor that these techniques entail, and unless the operation you are doing authorizes you by contract to use this type of techniques directly on the objective, **it should be considered to include actions to increase anonymity**, both at the level of the person and the audit machine, to increase the realism of the actions. Obviously, these anonymity techniques can also be combined with passive techniques for an extra level of caution that surely borders on the excessive, but avoids being discovered carelessly (for example, looking at an *Instagram* story). Let's see what techniques there are.

NOTE: In this course it is completely forbidden to use techniques that involve the "engagement" of any class with people, unless it is a simulation, and the other person is a partner who knows exactly what you are doing.

The "Intelligence Cycle"

The works where information is collected from an objective does **not only require collecting information and nothing else**, but this information must be processed in a cycle: once we obtain information about an objective it is possible that we can use it to obtain more associated objectives, or to deepen what we have found, so it is a process that is easily fed back. In addition, the process can be structured in five phases.

- **Preparation**: Which is a bit what we will talk about in the next section, preparing the environment to collect information, and identifying the best sources of information or tools
- **Collection**: What most of the document
- **Processing**: The information collected must be properly organized and ordered to facilitate its analysis
- **Analysis**: The information is interpreted, and all kinds of conclusions are drawn from it.
- **Presentation**: A report is made with the information found and its conclusions, which is the final objective of this whole activity.

The following image describes this process:

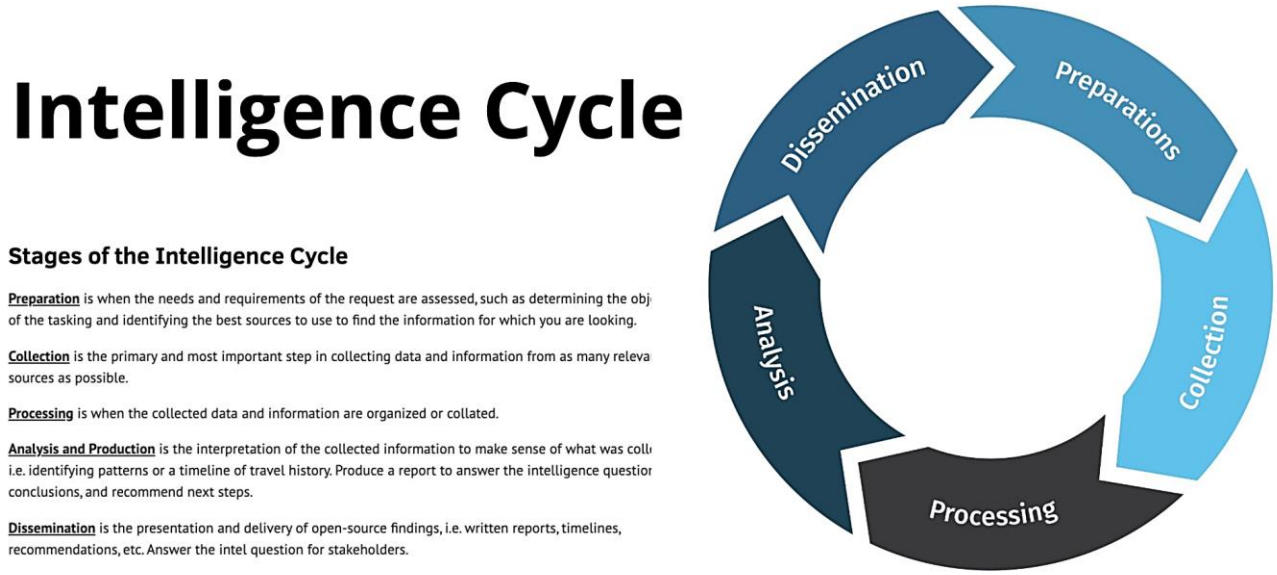


Figure 2. Intelligence cycle. Source: https://www.sans.org/blog/what-is-open-source-intelligence/

"Personal" Anonymity: Creating a False Online Identity

Use of temporary emails to register on services

One of the main reasons why we receive spam is for using our main accounts to register for services whose privacy policy leaves much to be desired, their customer database is stolen, or is directly designed so that we accept all kinds of communications by default (the fine print that we do not read!). Also, when registering for an online service we can leave the account visible by carelessness.

In offensive operations such as those described here, an account of this kind allows us to register inn some service or forum that the objective uses to study what he or she does without revealing any kind of information about ourselves. Obviously, **they should not be used to receive periodic notifications of anything because, as we will see, they are disposable**.

We can mitigate some of these problems by using **a disposable email account**. It is also applied in scenarios where we have to leave a contact person for a website, technical service, etc. but we do not want to expose the identity of a real person, thus creating an illusion that "deceives" the interlocutor. The register process in a RRSS normally does not support this type of email accounts, and it would not be convenient to use them if we want to keep them for a while (the recovery mail would disappear), so we should use a new account in a more standard provider in these cases, **obviously using data that does not link us with it**. In other words, we must always investigate with disposable emails that cannot be associated with our identity, what changes is the ease or speed with which they are discarded and the planned "lifetime" they must have 😊.

Disposable emails are a service that **allows us to receive emails at a temporary address** that **self-destructs** after a while or when we voluntarily decide to do so. They are especially applicable in forums, Wi-Fi portals (hotels, conferences ...), websites, and blogs that ask visitors to register before they can see the content, post comments, or download something. Although it seems incredible, if we do it with our real email in many cases we are opening our email address to be included in a spam distribution list, and therefore be "besieged" by emails of this type, in addition to being able to be identified. The same happens in any public email address that we leave on a website. With a disposable account, this is not possible: as soon as we "kill" the account there will be no trace of it.

Apart from the temporality of these addresses, there is another feature of these that we must consider, and that is that, although they have no problems receiving emails from other addresses, **sending emails from them is very limited**, fails a lot, or directly is a non-free service. However, to register on different sites it is only necessary to receive emails.

Email On Deck

Another additional problem of these services is that some sites do not allow you to use one of these addresses to register them, and that is why it is useful to know several 😊. Although there are several providers of this type of email, we will describe 3 to give you enough work material. The first one is ***Email On Deck*** (https://www.emailondeck.com/es/eod.php) that will immediately create a random name email account ending in a default domain and that we can use at the moment:

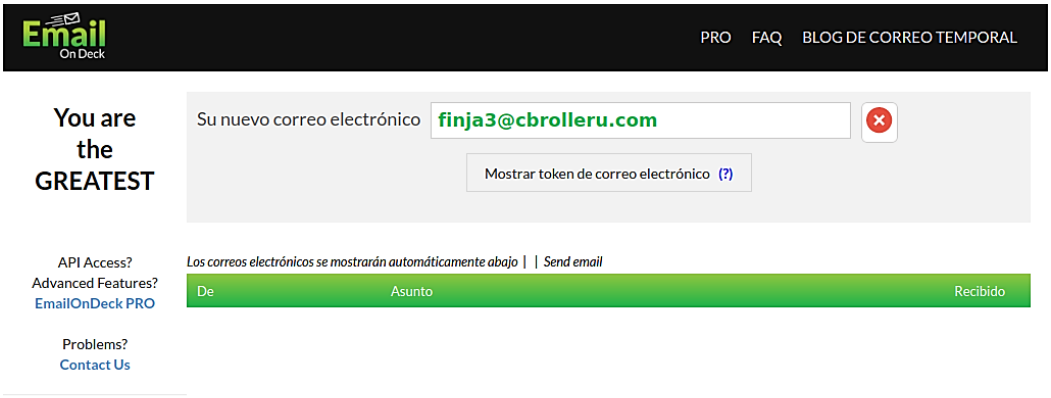


Figure 3. Newly created Email on Deck account

As we can see, you can receive emails from Uniovi without problems:

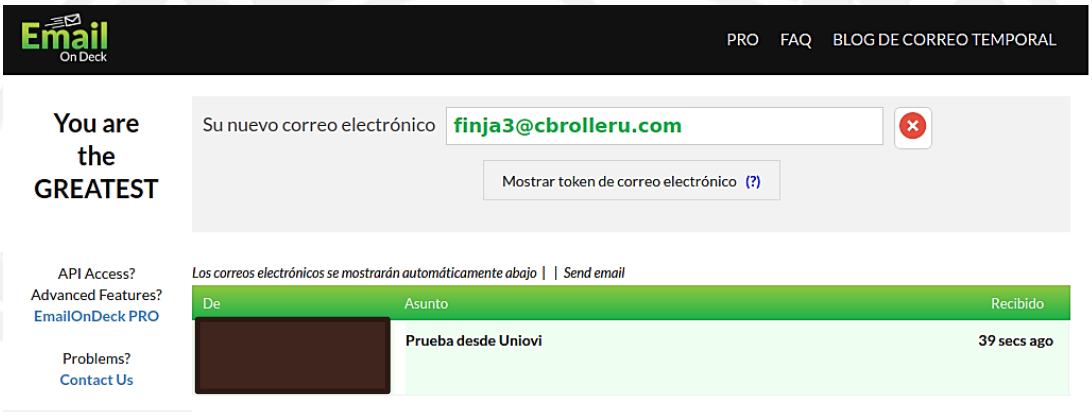


Figure 4. Email on Deck receives emails without problem

When we finish using this temporary address, we can press the delete button and there will be no trace of it:

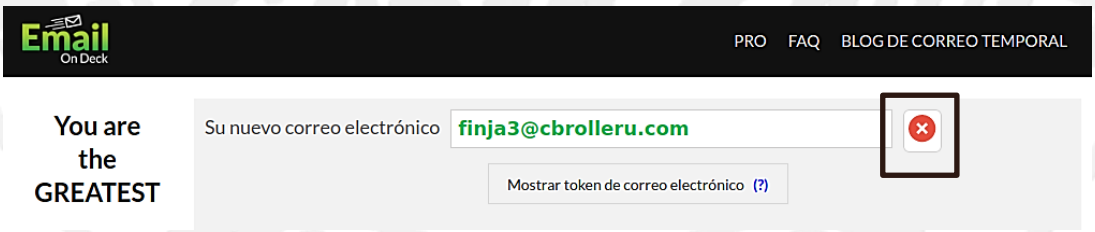


Figure 5. Remove all traces of us from Email On Deck

Our next access to the web will generate a new address randomly and there will be no trace of the emails received in the previous one, **which are unrecoverable**. Sending emails with this service requires a paid subscription, but we do not need it in this course. However, this opens the door to use it in "serious" research in the future if you decide to dedicate yourself to this 😉.

Figure 6. New *Email on Deck* account

GuerrillaMail

The next service that we are going to see is *Guerrillamail.com* (https://www.guerrillamail.com/es/), with similar characteristics to the previous one and that, in principle, also allows you **to send emails for free**, although the tests we made of this last option have not been satisfactory, at least with Uniovi. A priori it is not possible to know if it was a temporary problem of the service, a blocking / spam policy that Uniovi has (perhaps it works for other destinations) or for another reason. Again, simply by accessing the page we already have a temporary email that we can use:

Figure 7. Newly created account with *GuerrillaMail*

You can modify the name of the account from which to send an email at will and, as we see, the features for writing an email are the typical ones. As an added feature, it allows sending attachments (up to 150Mb). These emails and attachments will be deleted after one day and it should be noted that **they do not offer full privacy**, since the IP from which we send it will be part of the email headers to avoid being automatically classified as spam. However, if we do this from the ToR network (next section) so that our machine is also more anonymous... (*wink* *wink*). NOTE: Actually, using the ToR network in real research is always advisable just in case 😊.

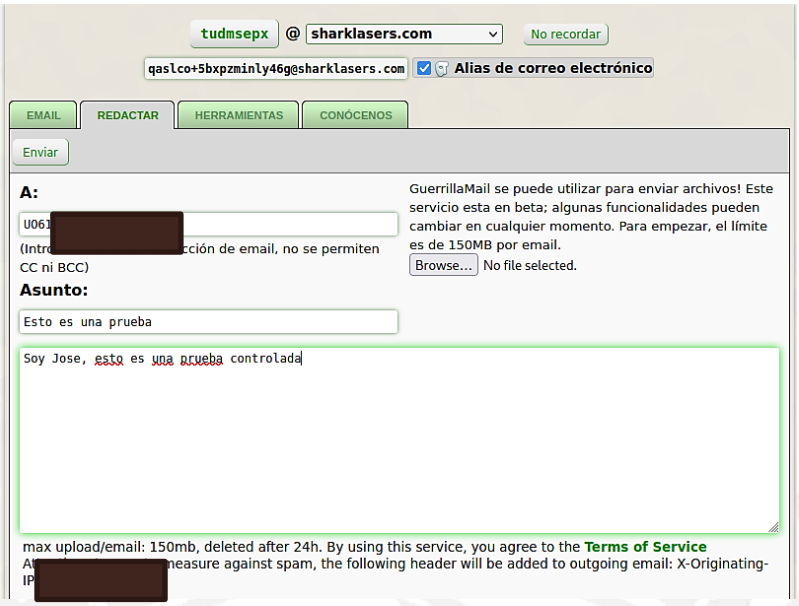


Figure 8. *GuerrillaMail* send email test

Another interesting feature is that we can change the domain from which the email originates very easily in case any of them is not supported by a service:

Figure 9. Domains available with *GuerrillaMail*

We finally see how the reception of emails from Uniovi happens without problem. Clicking on "Do not remember" we will destroy the email address and the emails received:

Figure 10. Receiving an email with *GuerrillaMail*

TempMail

Finally, a third service of equivalent capabilities is ***TempMail*** (https://temp-mail.org/es/), again only suitable for receiving emails.

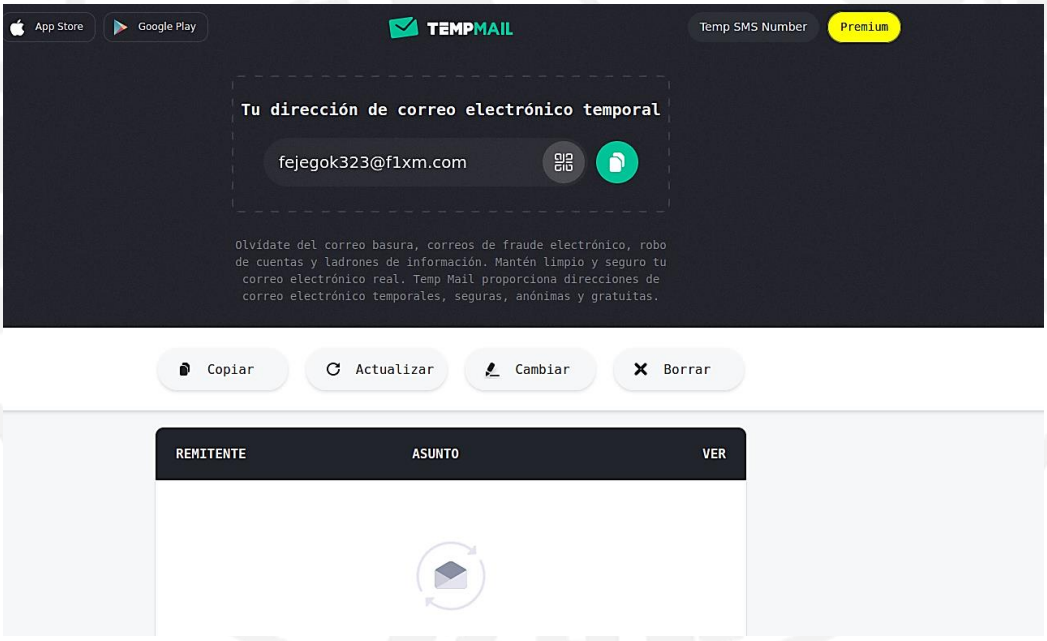


Figure 11. *TempMail* disposable mail service

We see how you also receive emails from Uniovi without any problem:

Figure 12 Receiving emails with *TempMail*

And again the "Delete" option immediately generates a new account:

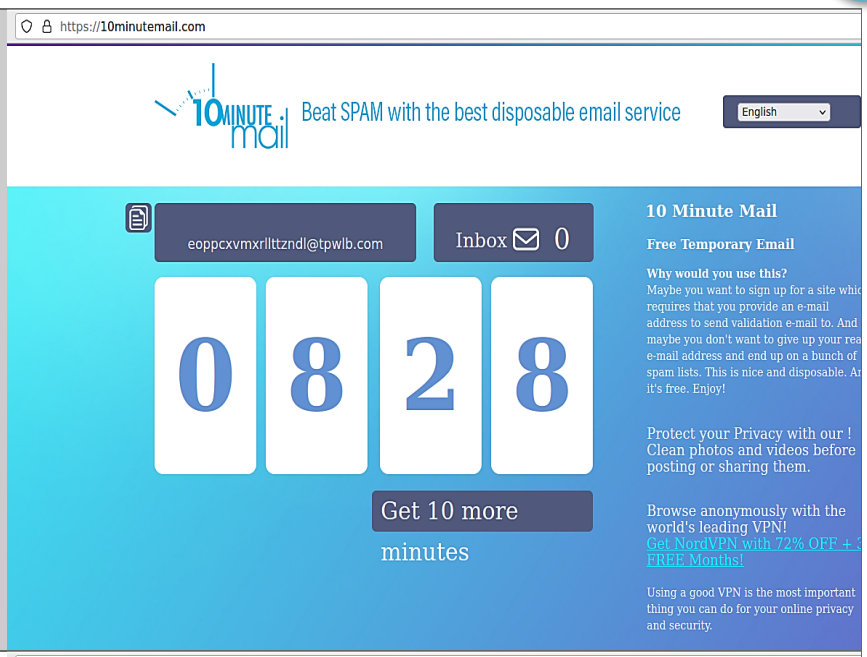
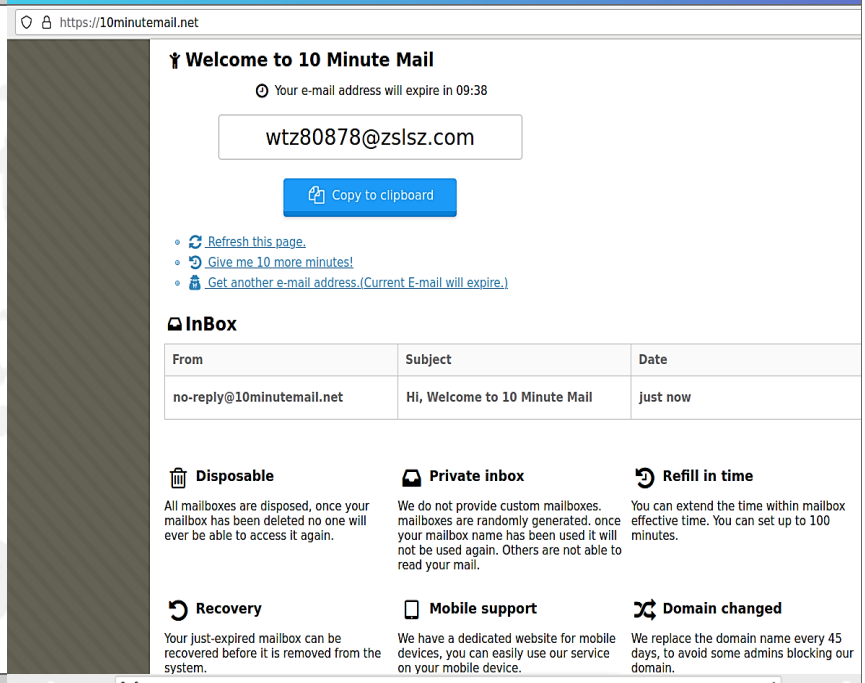
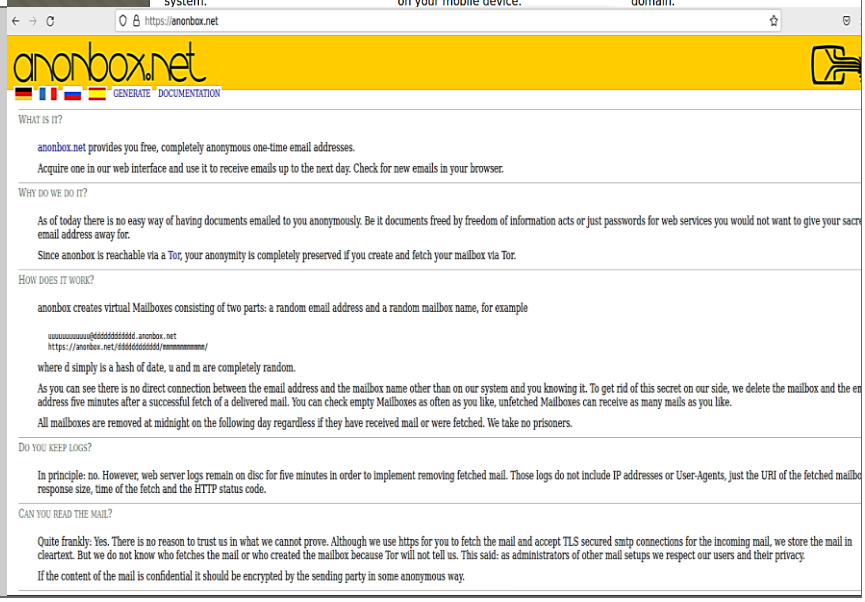
Figure 13. Deleting emails with *TempMail*

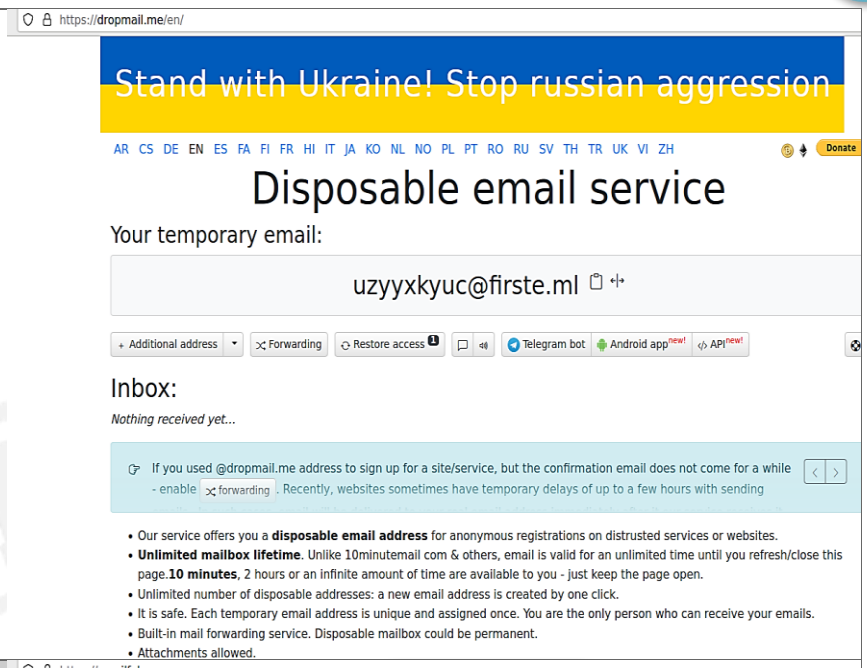
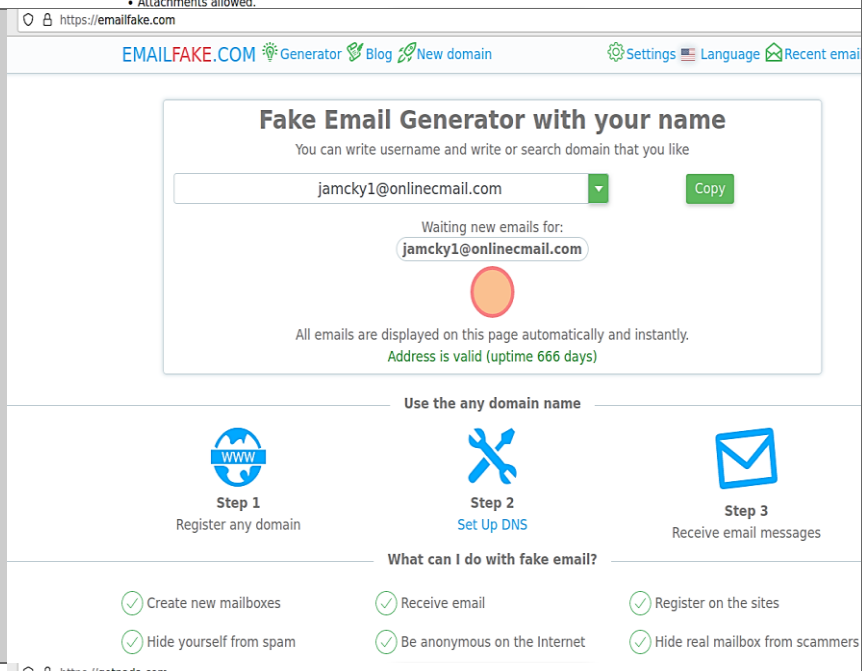
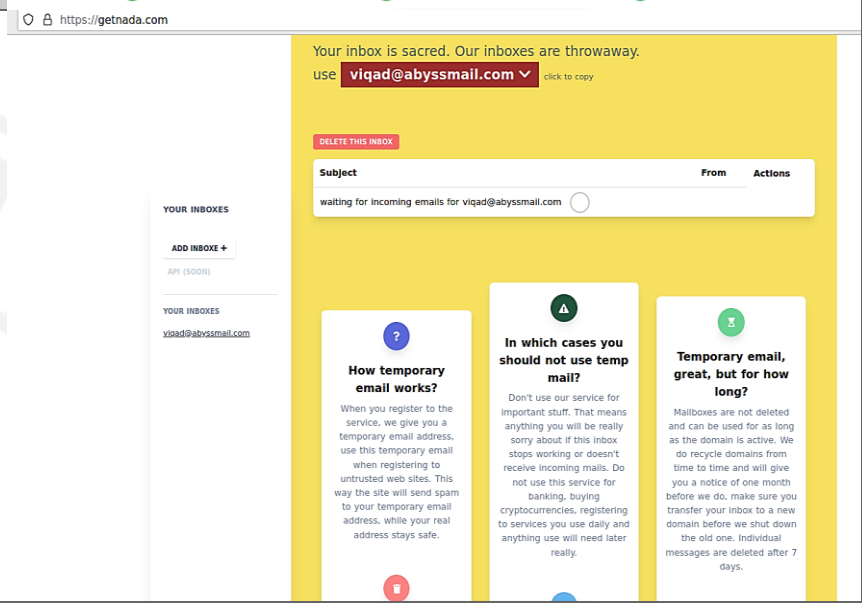
Other temporal mail providers

As we said, these temporary providers can be blocked by our target, or they can disappear after a while. That is why it is very important to have many collected to use them when necessary. It is also possible that some of them do support sending emails for free, even if it is with limitations. The following table lists a large number of providers equivalent to those seen in the previous subsections. All of them have a very similar operation to the ones we reviewed.

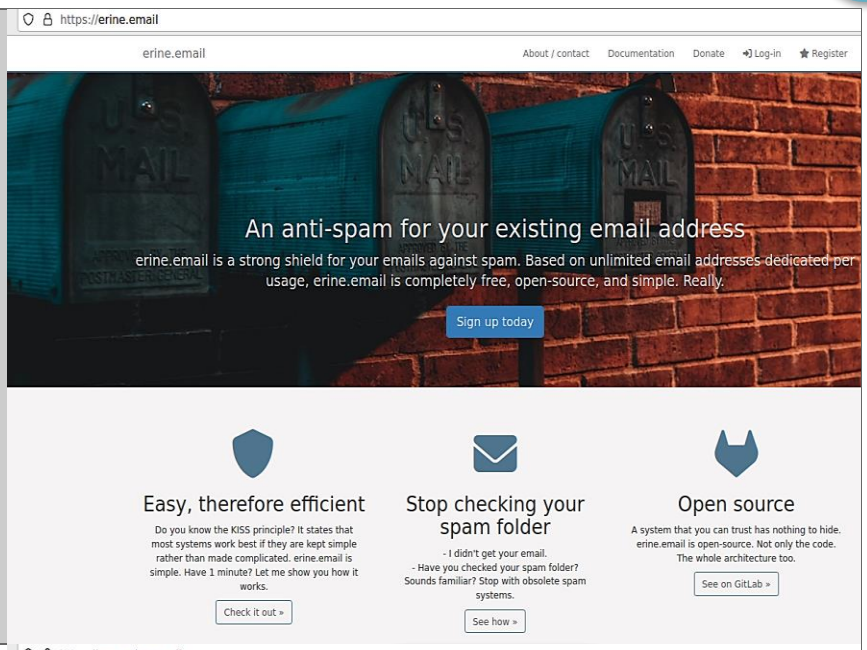
Table 1: Temp Email providers

Provider	Service's screenshot

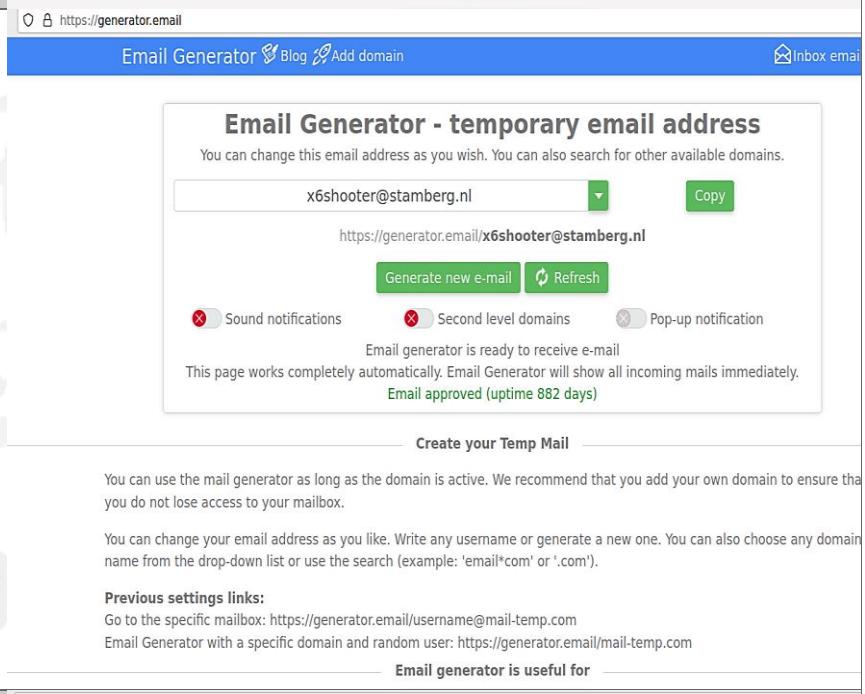
https://10minutemail.com	
https://10minutemail.net	
https://anonbox.net	

https://dropmail.me/en	
https://emailfake.com	
https://en.getairmail.com o https://getnada.com	

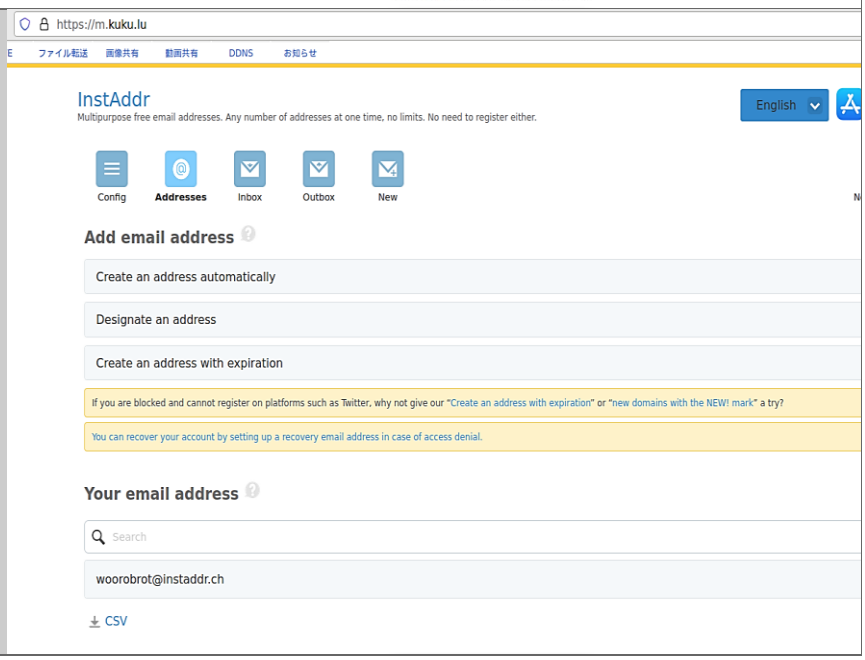
https://erine.email

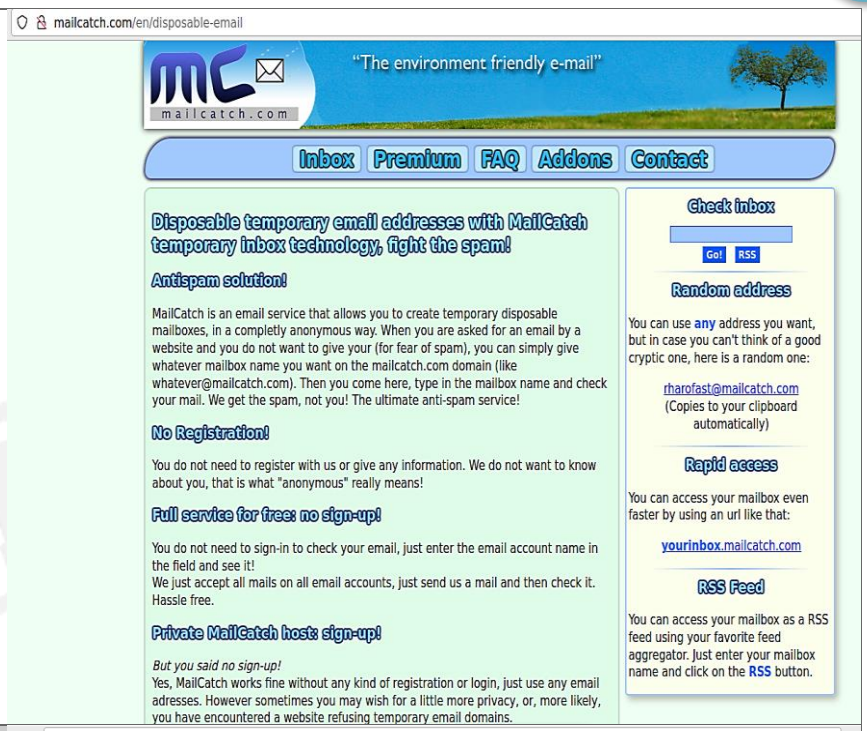
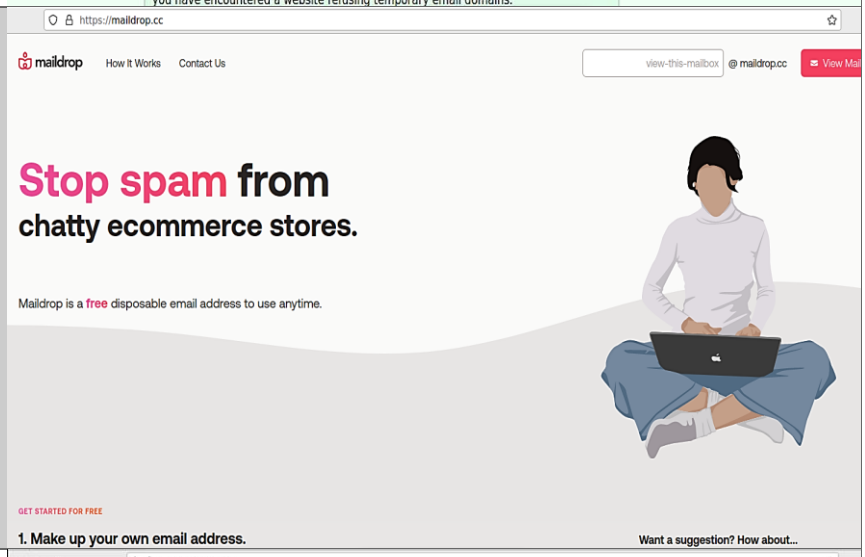
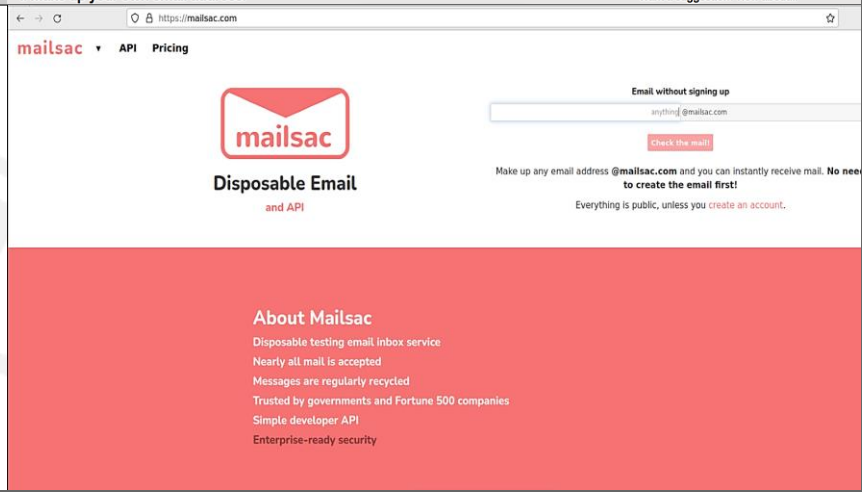


https://generator.email

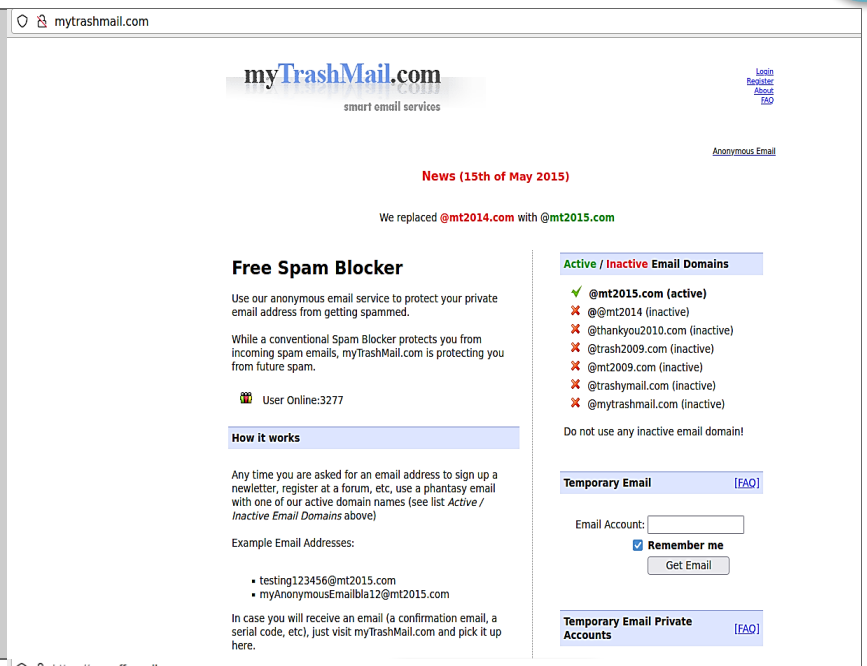


https://m.kuku.lu

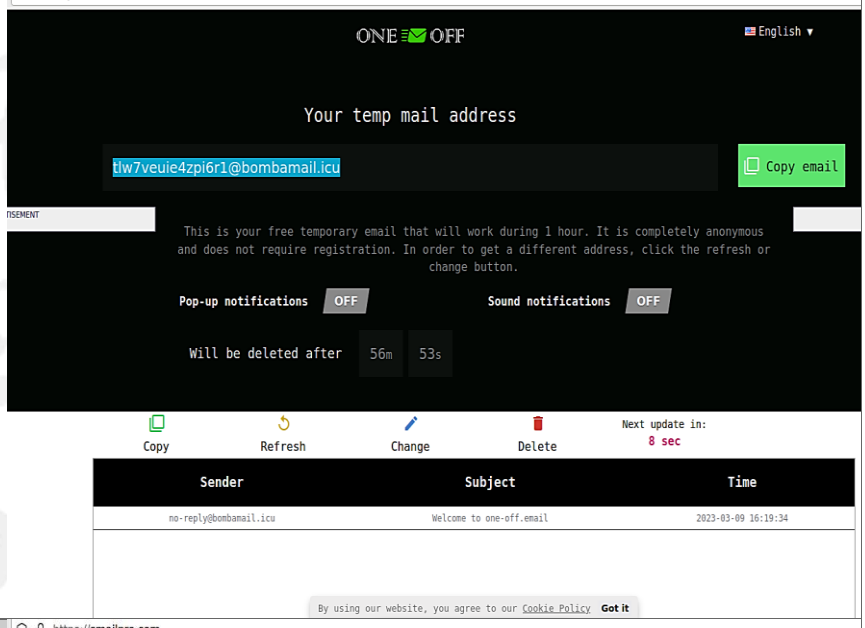


http://mailcatch.com/en/disposable-email	
https://maildrop.cc	
https://mailsac.com	

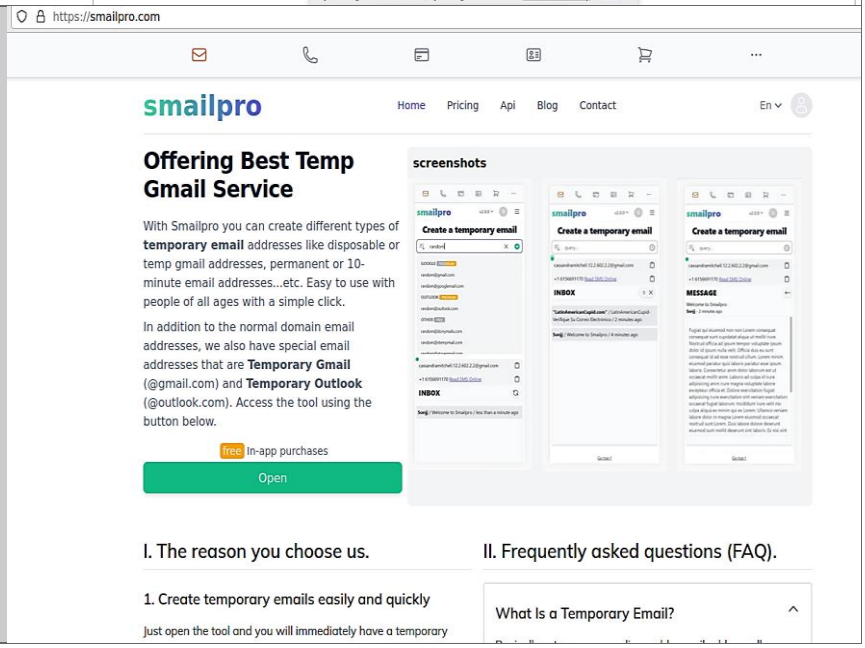
https://mytrashmail.com

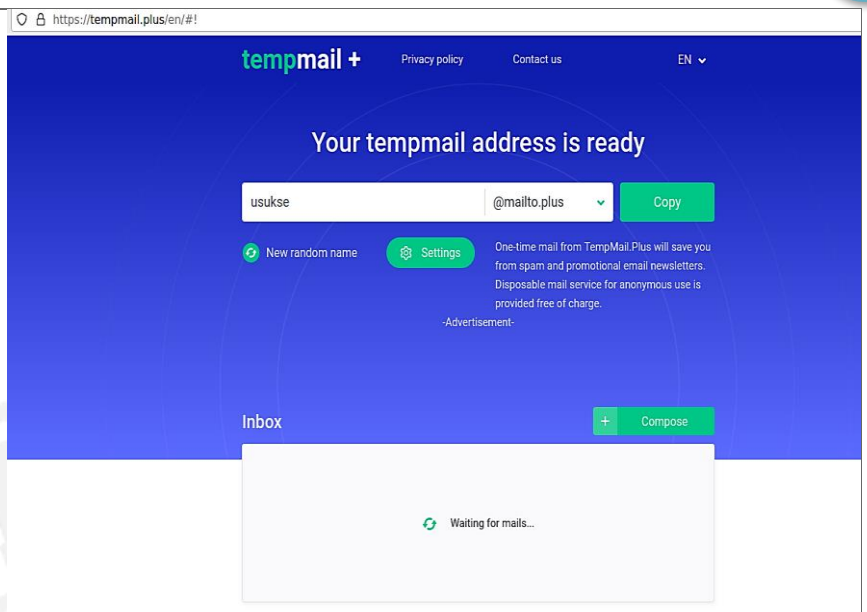
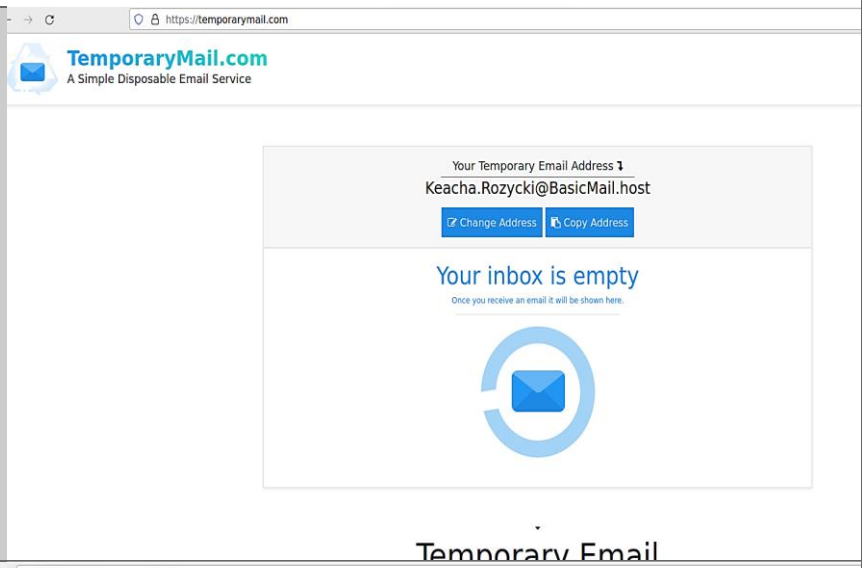
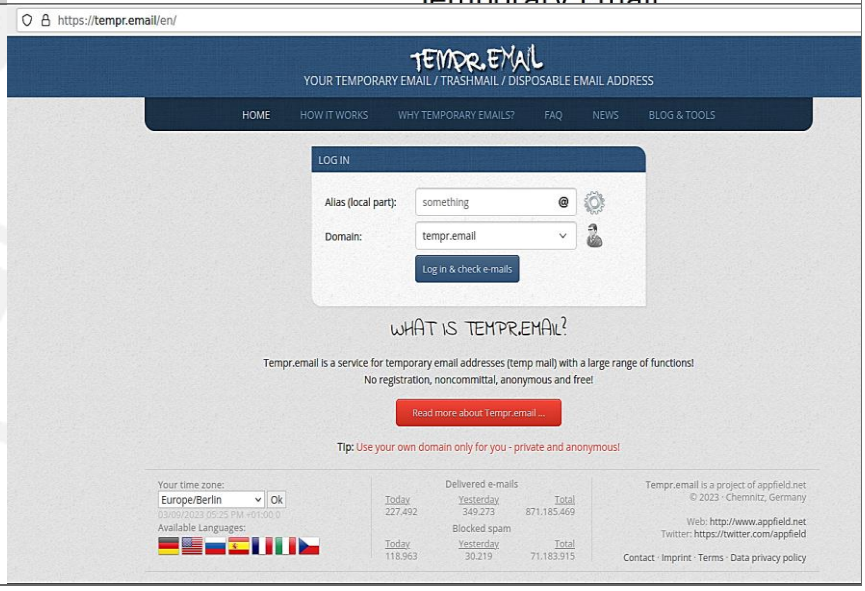


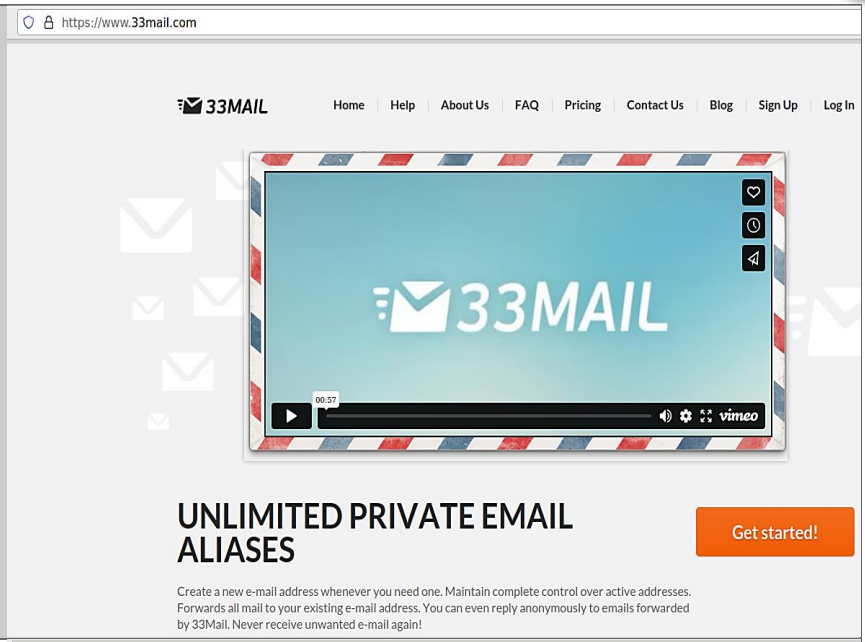
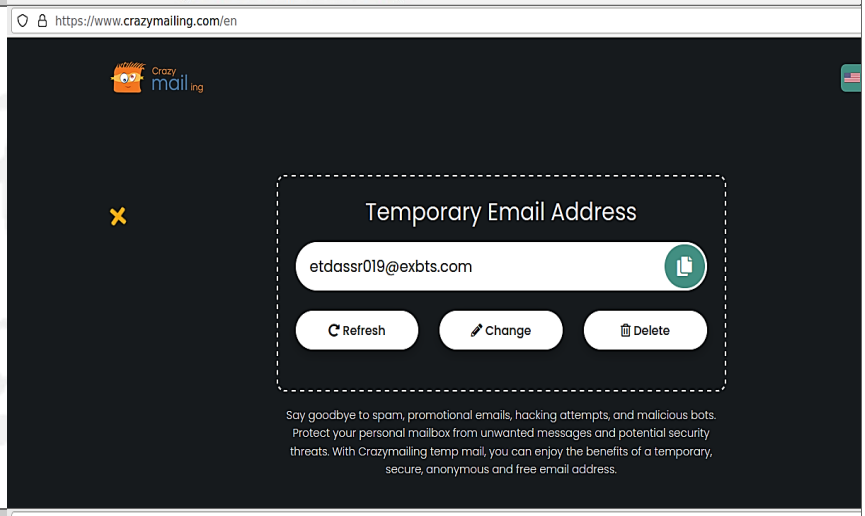
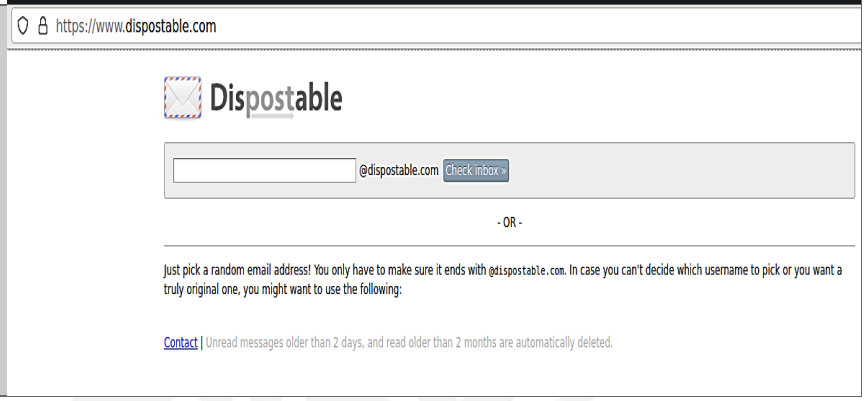
https://one-off.email/ru

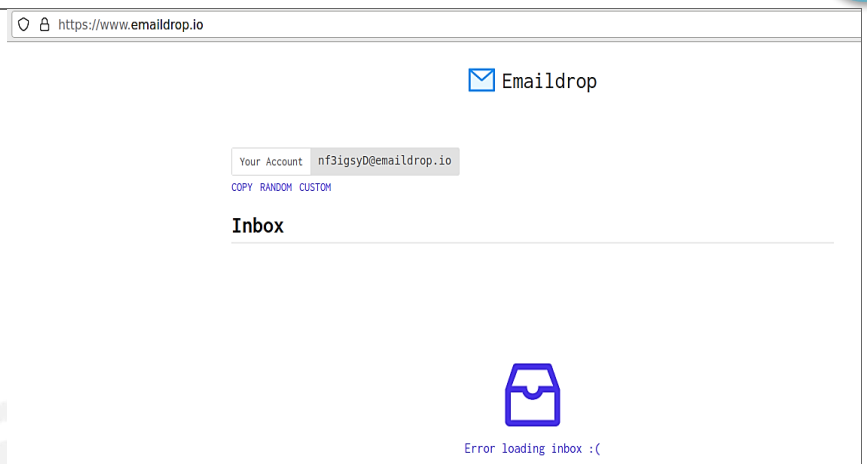
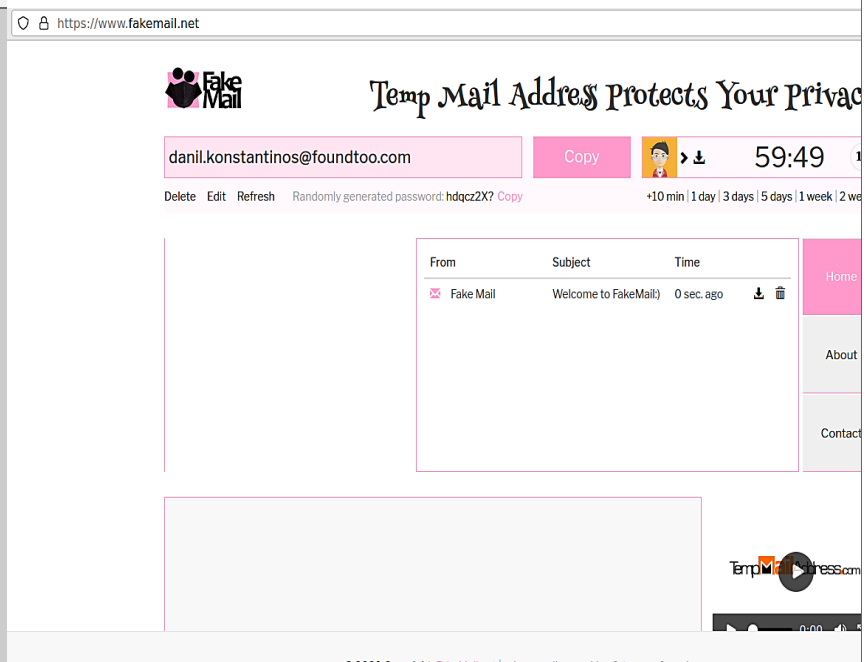
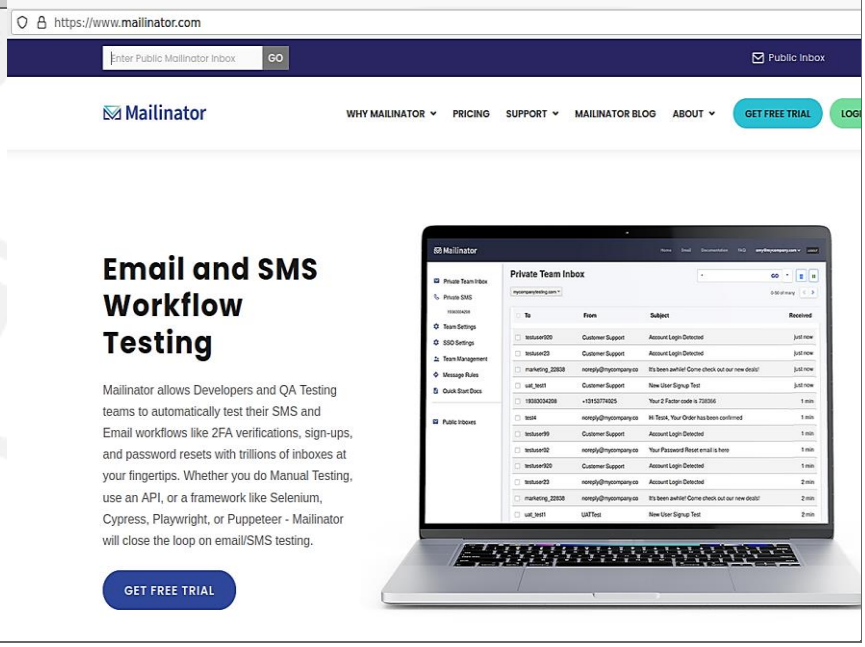


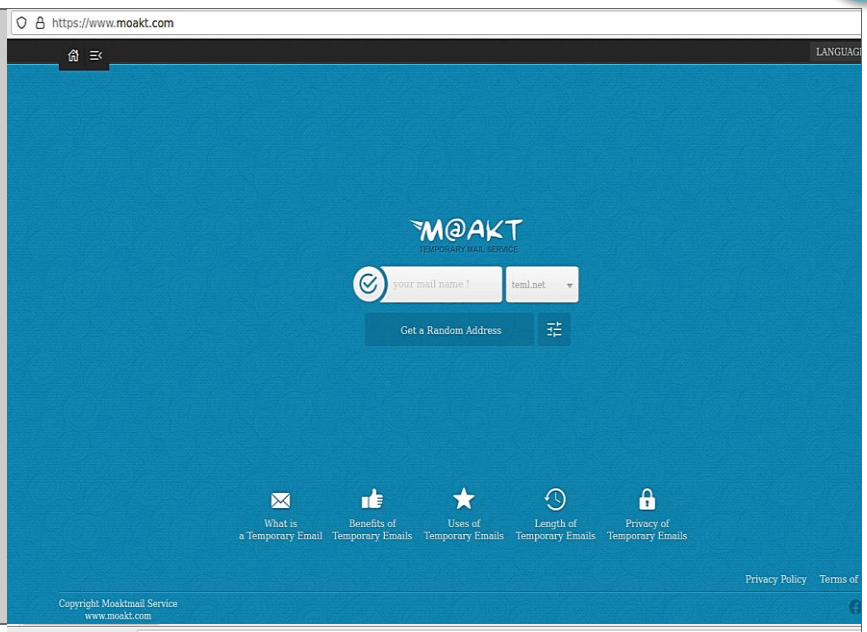
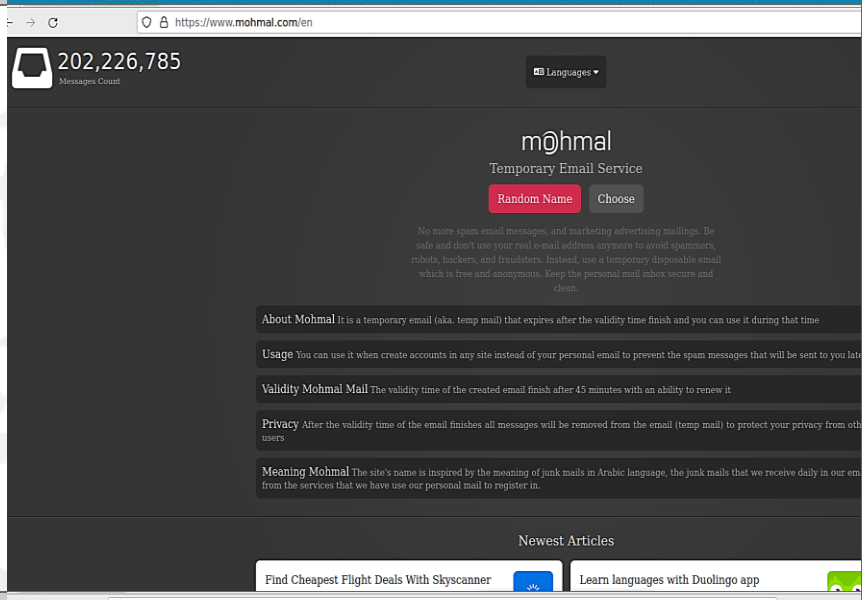
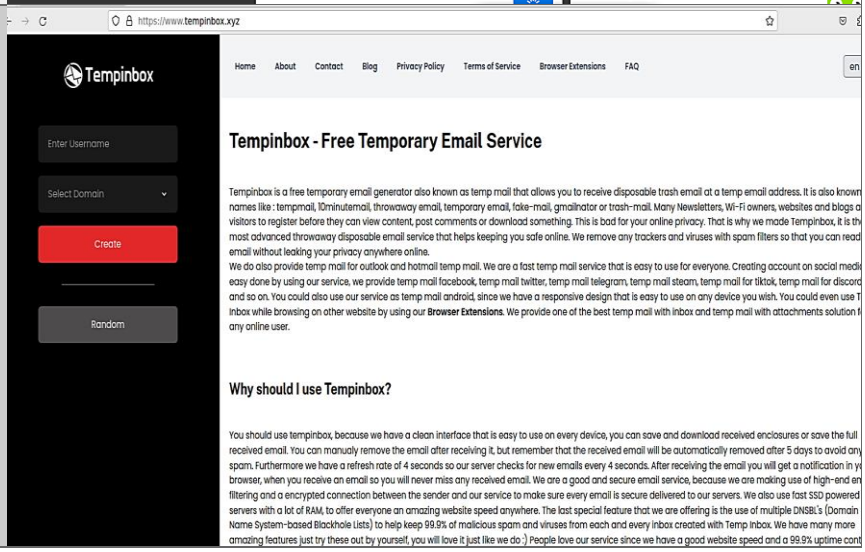
https://smailpro.com

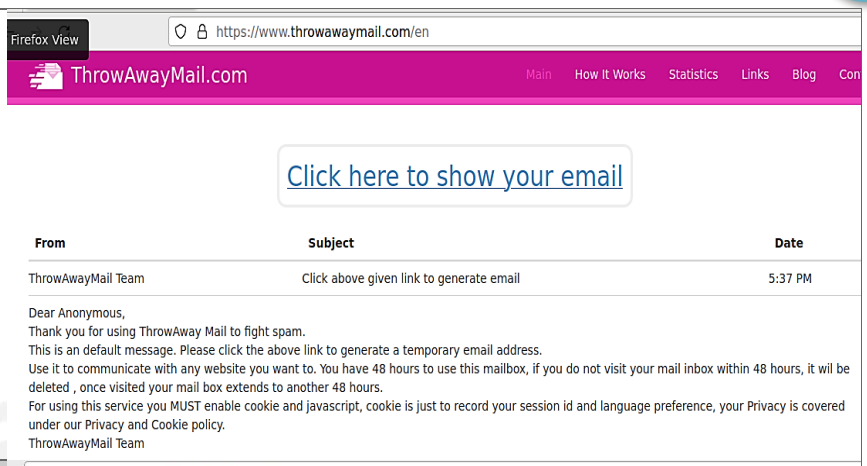
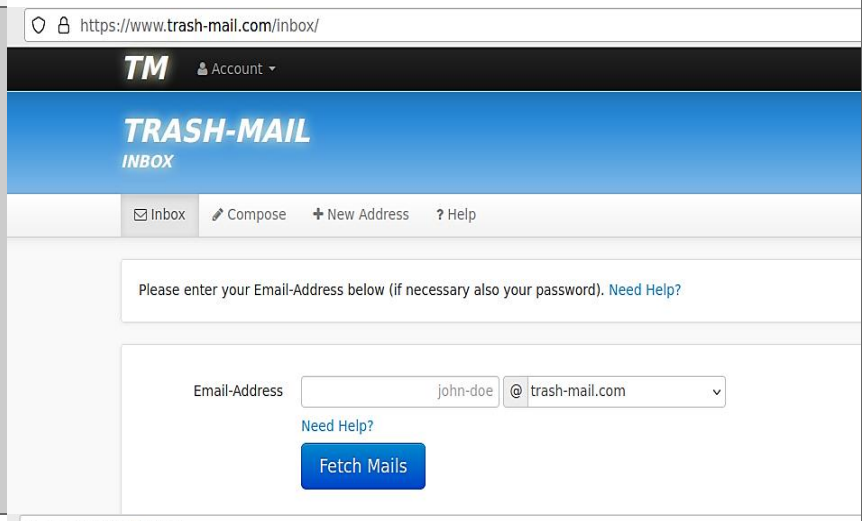
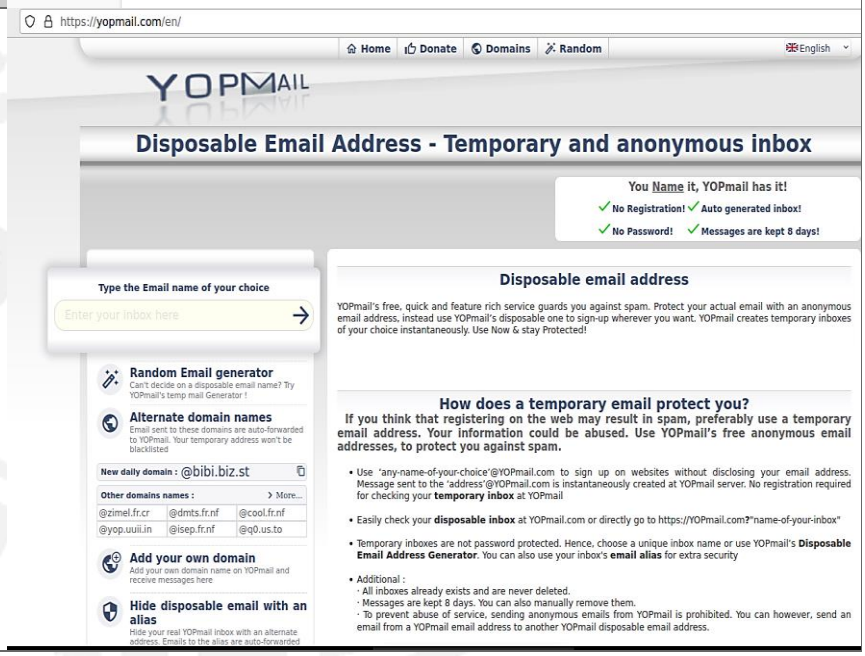


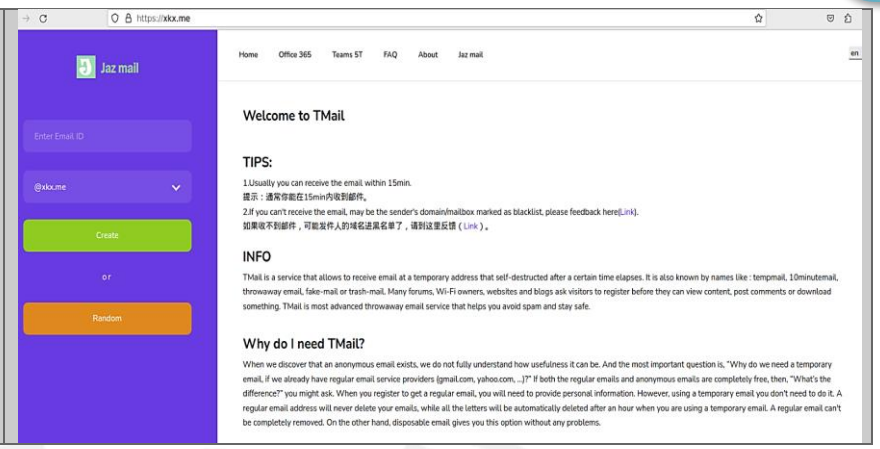
https://tempmail.plus	
https://temporarymail.com	
https://tempr.email/en	

https://www.33mail.com	
https://www.crazymailing.com/en	
https://www.dispostable.com/	

https://www.emaildrop.io	
https://www.fakemail.net	
https://www.mailinator.com	

https://www.moakt.com	
https://www.mohmal.com/en	
https://www.tempinbox.xyz	

https://www.throwawaymail.com/en	
https://www.trash-mail.com/inbox/	
https://www.yopmail.com/en	

https://xkx.me	

Creation of fake identities

Although in the previous activity we have seen that creation of temporal emails is a real possibility to cover ourselves against identity theft, personal data, protect our accounts against spam, or contact investigated targets so that the used email address cannot be traced back to us, sometimes it is necessary to go further and create a complete **fake online person**. This section is included to make tests that exclusively **do not involve interacting** with investigated objectives, such as registering on sites to observe people's behavior or posts, but without giving your own data. **Remember, in this course it is strictly forbidden to engage with strangers**.

Impersonating someone is a crime. **Being someone fake to deceive another person and obtain private data is a crime** unless you are a law enforcement agent and have authorization to do so by some kind of ongoing investigation. Actually, this section is detailed more for the opposite reason: you have many more options **to end up being a victim of criminals who do this**. So read this to be more prepared, and **to know when "something smells bad with this person" more easily** but telling it to you as if you were going to do it, which I think is easier to understand. And beware, this is not Hollywood. **This can happen to you in a high-level spear phishing attack or an elaborate scam**. If one day you have a position of responsibility, you can be a target more easily than you think. So be careful.

What does this mean? That there are services that provide us data that allows to create a complete false identity, so we can use it to register on other services when we do not want to expose our real identity. **While this is illegal or not allowed by some platforms or services** (for example, some social networks prohibit in their terms and conditions to do so), **it is a real option in many others** (that is, it does anyway, until they discover and remove it). In short: **any personal data can be falsified with a high level of credibility**.

One of the first things that is faked is the **profile picture**. **Stealing someone's identity is a crime**. Placing a drawing or avatar can also be **an infringement** (if we use a copyrighted one, but we have not paid for it) or **not be allowed by the service** (which requires a realistic person's photo). Some services to create false identities already give us one (we will see that later), but it can be easily noticed or found out that it is a "stock" photo (Ex: searching on *Google*). For these cases, there is a service called https://thispersondoesnotexist.com/ that generates us using AI algorithms a portrait photo of a person that does not exist, but with incredible realism (includes different gestures, backgrounds, ages, sexes, races, etc.). Every time we reload the page we will generate a new random photo that will not be repeated, **so if we like any of them we must save it**. All the photos in the following figure are completely false, and we can see how many of them are very realistic:

Figure 14. No, none of these people exist ;)

Obviously, we must choose one that fits with the information we give of our false identity (sex, race, physical characteristics ...). Another advantage of using this type of service is that searching for the profile **picture with *Google's* reverse image search** will rarely give any results, since, as we said, **these photos are generated randomly**.

However, while this covers the profile picture section, **we still need the rest of the data** to create a "fake person". **In a real audit**, it is important to have this data predefined to give them consistently in different profiles that we create, so that our "cover" is not discovered by an inconsistency. This is not trivial. If we are going to investigate someone, it is possible that that person has technical knowledge, and we have to be very professional so that we are not discovered:

- The investigated person **can try to corroborate our history** by contacting third parties supposedly associated with us (which we ourselves will have created, because when a person is created, his environment should be also created if we get serious or a high realism is required).
    - Creating a set of fake people who form a "brotherhood" or associated group is not far-fetched. We want to control the interactions that can "betray" us, but it can also be used to extract even more information. One of those fake people may successfully manage to get more information from the investigated person if the "plot" is tick enough. This document is not called "human hacking" (just) for making a joke 😊.
    - Think about yourself: *what do you trust the most? A profile with friends or in another without them?* The answer is obvious, and you know it. Put yourself in the shoes of the investigated person and act accordingly. It is a person, just like you.
    - If you are creating someone fake who is part of a known environment or context, be prepared for questions like "*Do you know X?*", "*Have you been to Y?*" or "*What do you think of what happened in T (past tense)?*". In a real investigation you have to "do your homework" via extended OSINT research with the target environment (not in this course, I only tell you this so you can see how things are going in more serious contexts where the victim may be you).
- **We must always assume that the investigated person is someone very skilled and intelligent** and can try to actively look for "cracks" in our character. For example, asking for the same personal data at various points in time and looking for inconsistencies. To avoid inconsistencies, it is necessary to design the "play", its characters, and its profession well. Obviously, this minimizes the option of being inconsistent by accident as well. I can also tell you that it is something that requires practice.

- Get the idea that **you are playing a role-playing game, or you are in a theatrical play**. Your audience or peers need to dive into the "fantasy world" you have created so you can get the data you need. If not, you will fail, and all the effort will be useless.
- **Fake identities are "burned"**. Do not reuse fake entities. The world is very small, more than you imagine, and it is very sad to fail because someone has previously known your fake identity for a friend. The services I give you generate everything random. Do not worry, they are not going to be exhausted 😊.
- You can mix random information with parts of real profiles to give it more consistency if you want, **but never rely too much on someone** or it will be nonsense and you could incur in an impersonation crime. You have to keep being false. 100% credible, but false.
- It is also easy to get carried away by the ego in these cases. **You are a professional and you have been hired to extract information from someone to do your job**. Using these powers for something else makes you not a professional and easily leads you to be a criminal. Do not do it. Your job is to discover information to prevent misfortunes, **not to become the misfortune**. In this course **we do NOT want criminals**.

One service that allows us to do this is ***Fake Person Generator***: https://www.fakepersongenerator.com/Index/generate. With it we can generate a person of any gender, age, and state (it only allows to generate people residing in the US, although obviously then we can modify any data manually to "personalize" our identity). This service already gives us a profile image that we can change for one of the ones we saw before if we want.

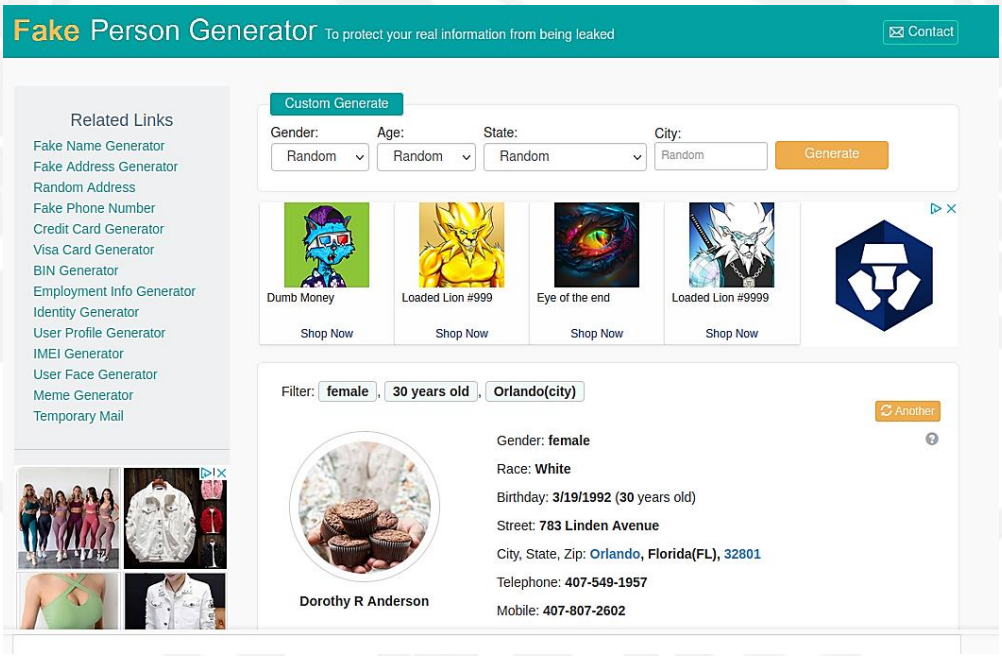
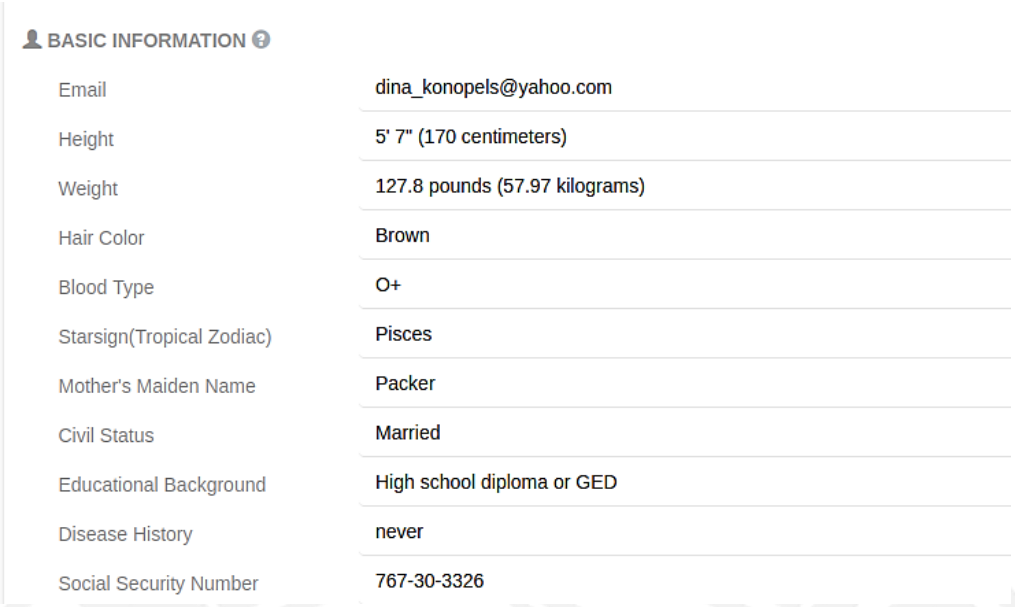


Figure 15. Profile created by *Fake person generator* (I)

We see how a series of basic data of the person are randomly generated (all false and random, obviously, but that comply with the format of a real data).

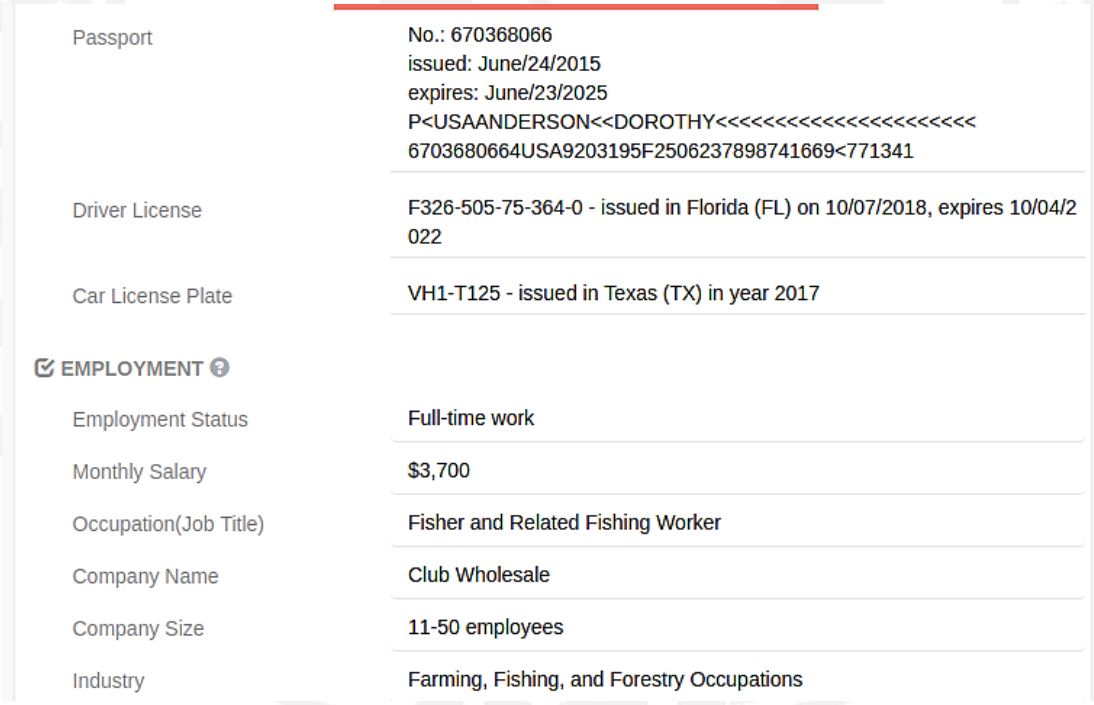


BASIC INFORMATION

Email	dina_konopels@yahoo.com
Height	5' 7" (170 centimeters)
Weight	127.8 pounds (57.97 kilograms)
Hair Color	Brown
Blood Type	O+
Starsign(Tropical Zodiac)	Pisces
Mother's Maiden Name	Packer
Civil Status	Married
Educational Background	High school diploma or GED
Disease History	never
Social Security Number	767-30-3326

Figure 16. Profile created by *Fake person generator* (II)

The page does not stop there, but generates false passport, driver's license, and car registration data, as well as current employment and finances. Keep in mind that, as we said before, an interlocutor can request this information to check that he is not in front of a bot, for example (and he is not, you are just a person, but fake ... 😊):



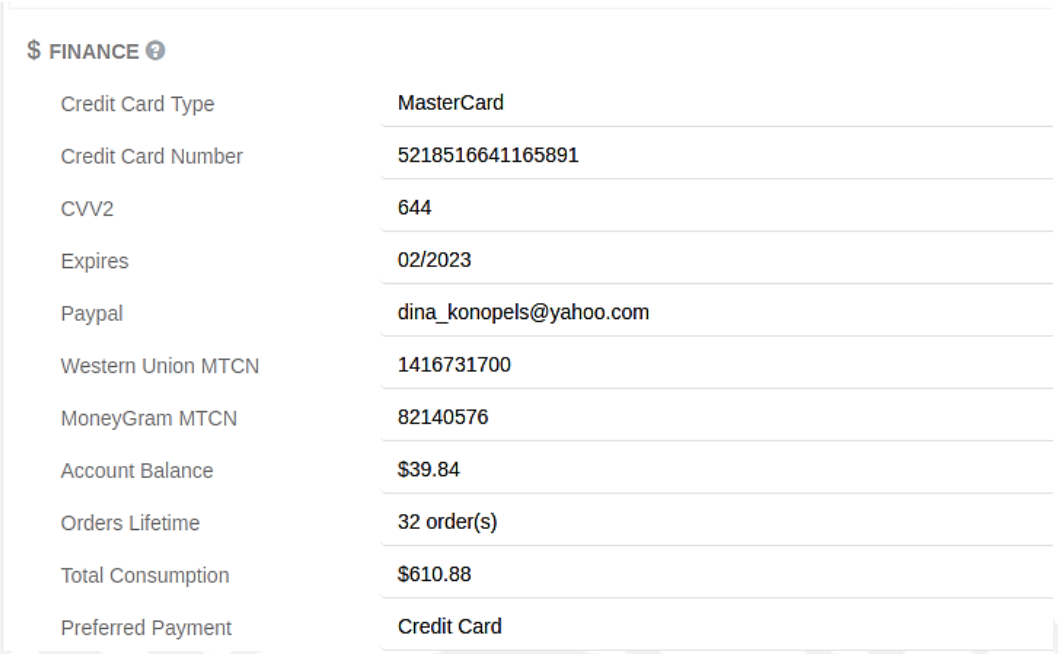
Passport	No.: 670368066 issued: June/24/2015 expires: June/23/2025 P<USAANDERSON<<DOROTHY<<<<<<<<<<<<<<<<<<<<<<< 6703680664USA9203195F2506237898741669<771341
Driver License	F326-505-75-364-0 - issued in Florida (FL) on 10/07/2018, expires 10/04/2022
Car License Plate	VH1-T125 - issued in Texas (TX) in year 2017

EMPLOYMENT

Employment Status	Full-time work
Monthly Salary	$3,700
Occupation(Job Title)	Fisher and Related Fishing Worker
Company Name	Club Wholesale
Company Size	11-50 employees
Industry	Farming, Fishing, and Forestry Occupations

Figure 17. Profile created by *Fake person generator* (III)

You can also see complete (false) financial data that is provided to be used in more specialized "covers" where you have to maintain a consistent story with data of this type.

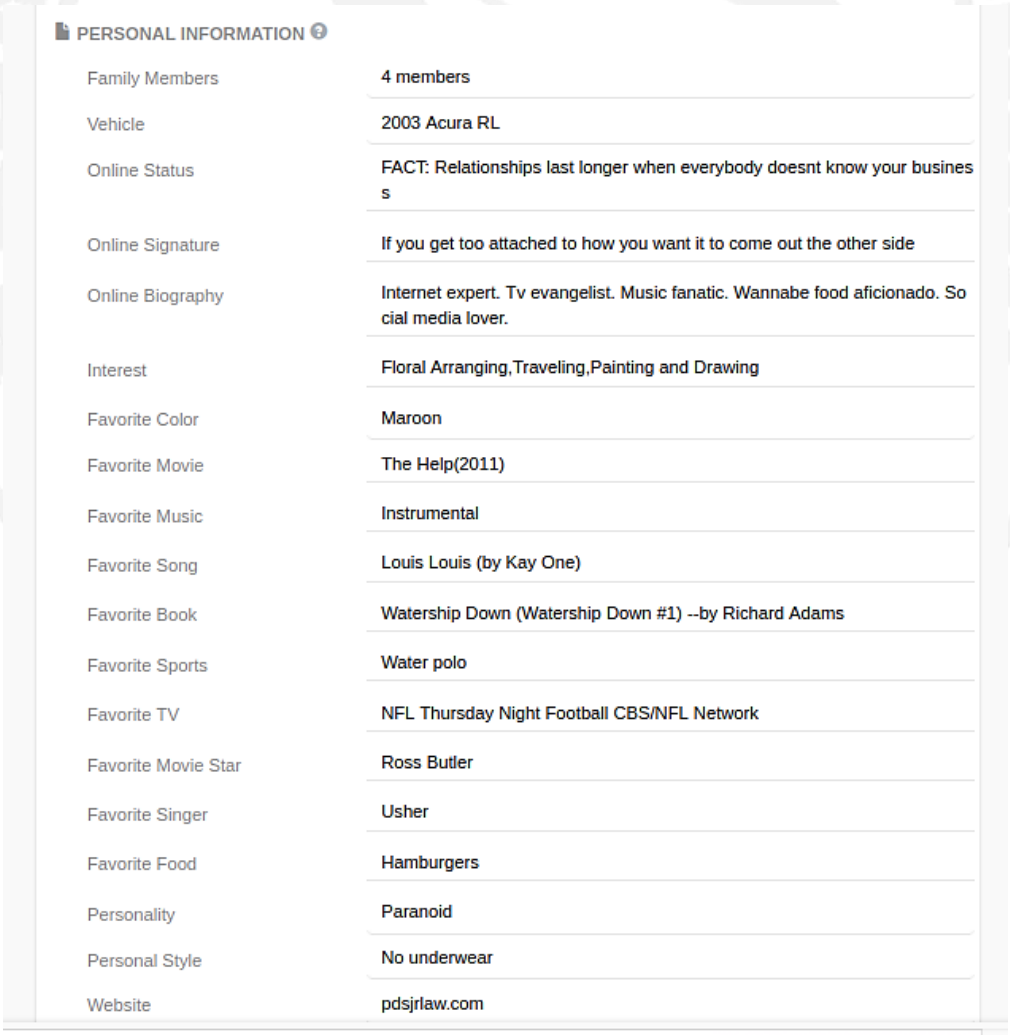


$ FINANCE

Credit Card Type	MasterCard
Credit Card Number	5218516641165891
CVV2	644
Expires	02/2023
Paypal	dina_konopels@yahoo.com
Western Union MTCN	1416731700
MoneyGram MTCN	82140576
Account Balance	$39.84
Orders Lifetime	32 order(s)
Total Consumption	$610.88
Preferred Payment	Credit Card

Figure 18. Profile created by *Fake person generator* (IV)

Additionally, personal data, family, hobbies, tastes, etc. usable in certain environments where this is applicable are included. Thus, we can create a much more complete "cover", since it really contemplates almost any typical social aspect that can be commented or asked in a network.

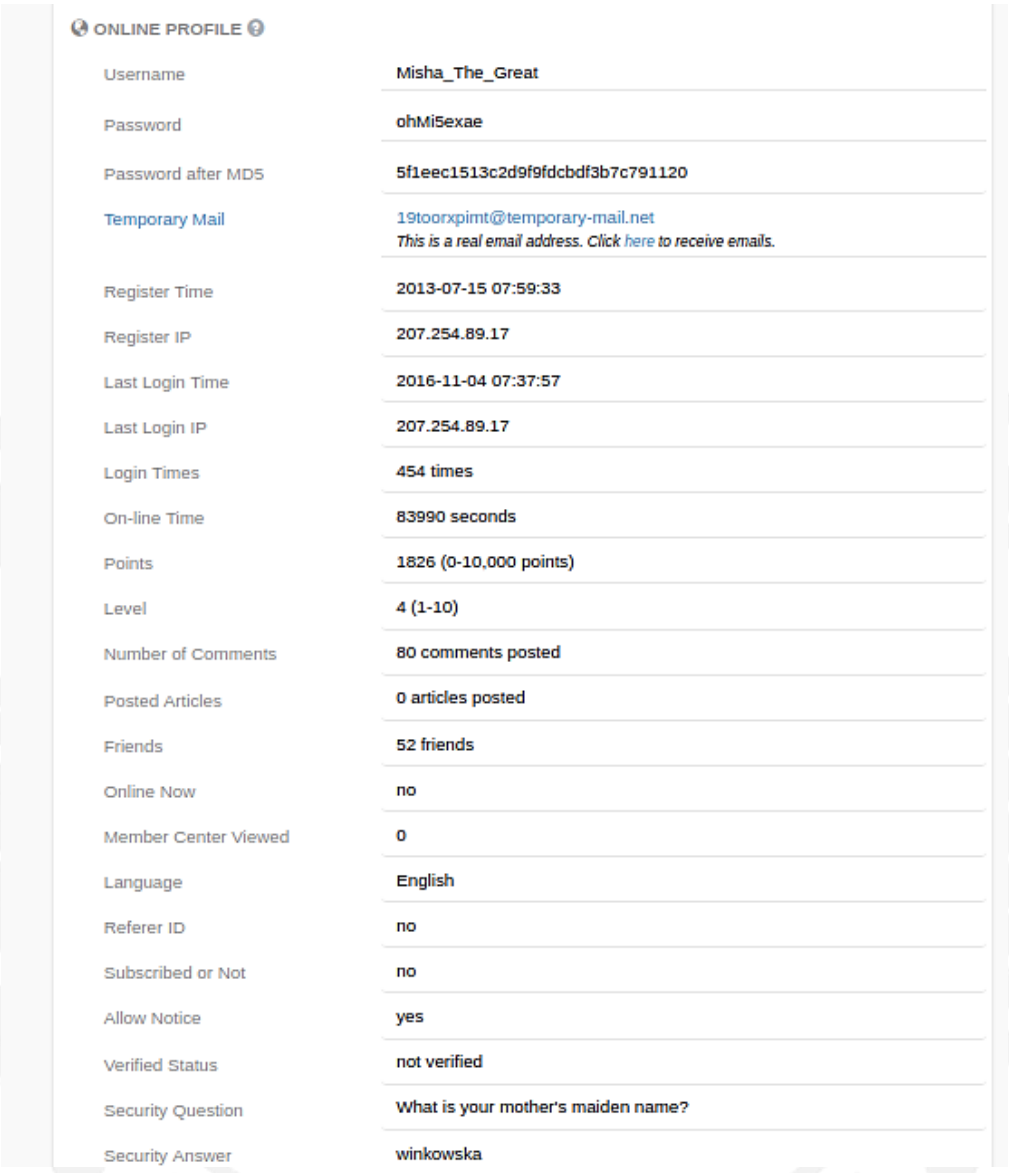


PERSONAL INFORMATION

Family Members	4 members
Vehicle	2003 Acura RL
Online Status	FACT: Relationships last longer when everybody doesnt know your business
Online Signature	If you get too attached to how you want it to come out the other side
Online Biography	Internet expert. Tv evangelist. Music fanatic. Wannabe food aficionado. Social media lover.
Interest	Floral Arranging,Traveling,Painting and Drawing
Favorite Color	Maroon
Favorite Movie	The Help(2011)
Favorite Music	Instrumental
Favorite Song	Louis Louis (by Kay One)
Favorite Book	Watership Down (Watership Down #1) --by Richard Adams
Favorite Sports	Water polo
Favorite TV	NFL Thursday Night Football CBS/NFL Network
Favorite Movie Star	Ross Butler
Favorite Singer	Usher
Favorite Food	Hamburgers
Personality	Paranoid
Personal Style	No underwear
Website	pdsjrlaw.com

Figure 19. Profile created by *Fake person generator* (V)

Finally, much more technical data is also generated, that can be used to register on services (includes a temporary email account, different to the ones we saw), artificially register the user in a service under testing, and a long etcetera. The service therefore focuses on randomly generating a large amount of different information, **and we are the ones who decide how to use it in the context of application**.



ONLINE PROFILE

Username	Misha_The_Great
Password	ohMi5exae
Password after MD5	5f1eec1513c2d9f9fdcbdf3b7c791120
Temporary Mail	19toorxpimt@temporary-mail.net *This is a real email address. Click here to receive emails.*
Register Time	2013-07-15 07:59:33
Register IP	207.254.89.17
Last Login Time	2016-11-04 07:37:57
Last Login IP	207.254.89.17
Login Times	454 times
On-line Time	83990 seconds
Points	1826 (0-10,000 points)
Level	4 (1-10)
Number of Comments	80 comments posted
Posted Articles	0 articles posted
Friends	52 friends
Online Now	no
Member Center Viewed	0
Language	English
Referer ID	no
Subscribed or Not	no
Allow Notice	yes
Verified Status	not verified
Security Question	What is your mother's maiden name?
Security Answer	winkowska

Figure 20. Profile created by *Fake person generator* (VI)

This service, although extremely powerful, has the problem that it only generates US identities. Another option is ***Fake Name Generator*** (https://es.fakenamegenerator.com) that, although it is not so powerful, and many of the names it generates (at least in Spanish) are somewhat "debatable", it does allow us to have a more appropriate identity for citizens of different countries (including **Spain**), whose data can be completed with those obtained in the previous service.

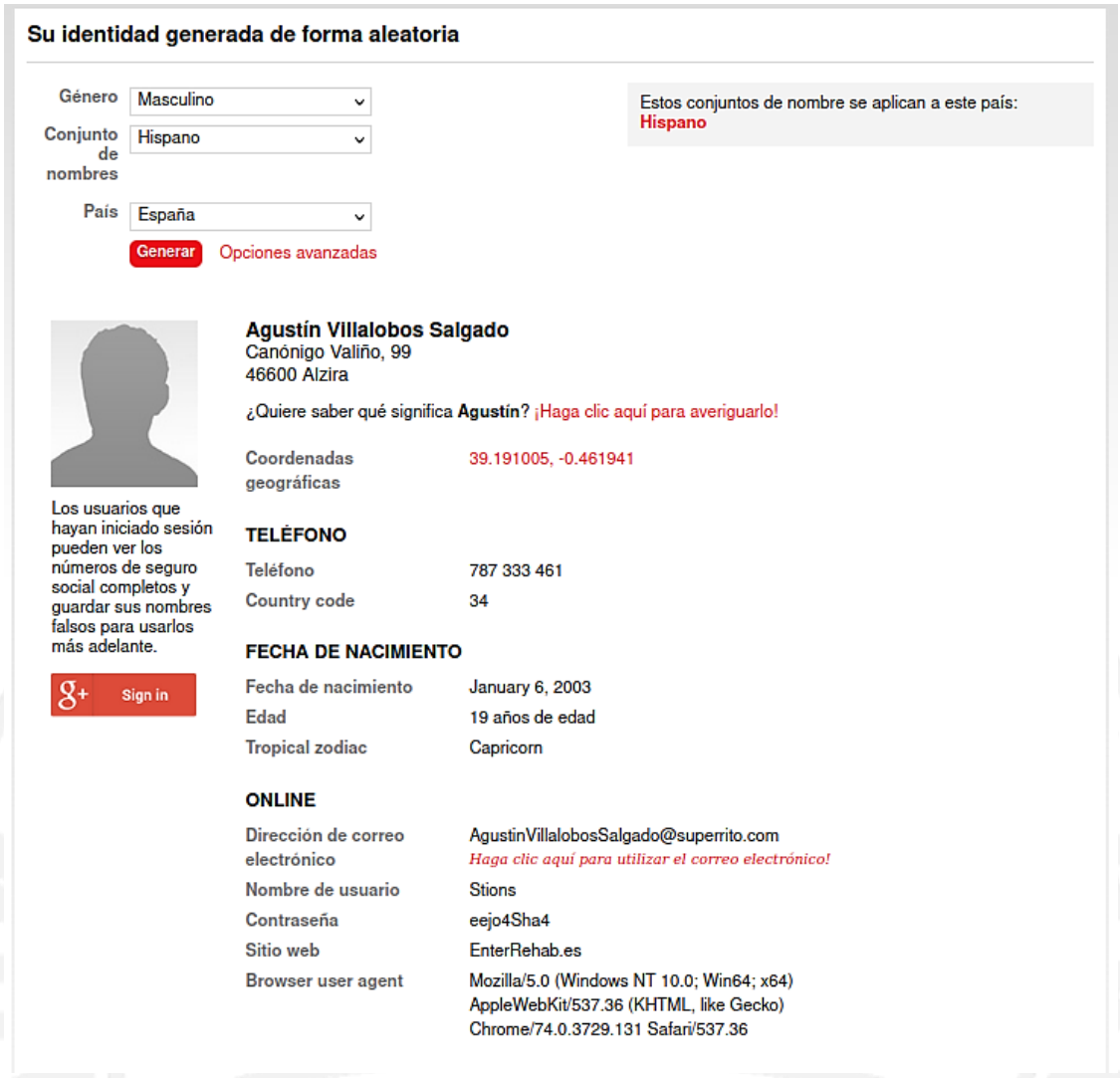


Figure 21. Profile created by *Fake name generator* (I)

However, as we can see, the false data it generates **is much less** than that of the previous service (although it includes a real temporary email account):

FINANCE

MasterCard	5134 4656 8731 2465
Vence	6/2027
CVC2	777

EMPLOYMENT

Empresa	Gart Sports
Ocupación	Aircraft pilot

PHYSICAL CHARACTERISTICS

Altura	5' 7" (170 centímetros)
Peso	164.8 libras (74.9 kilos
Tipo sanguíneo	B+

TRACKING NUMBERS

Número de seguimiento de UPS	1Z 4V6 59E 41 1184 740 8
Western Union MTCN	4671732598
MoneyGram MTCN	86202379

OTHER

Color favorito	Azul
Vehículo	2006 Lexus RX
GUID	88341456-78bc-4b52-ae61-f0bca5c8d255
Código QR	Haga clic aquí para ver el código QR de esta identidad

Figure 22. Profile created by *Fake name generator* (II)

In short, using one of them, **a combination of both**, or, as we said, a "story" that we create based on a combination of false data that we provide, plus the data provided by these services, together with an email account (temporary or not) not associated with our person, **we can be practically who we want on the Internet** and, if we maintain consistency in our stories and do not reveal information that gives us away, that turns out to be 100% credible in the eyes of any standard user. Interacting with someone for active OSINT tasks can therefore be viable from a completely anonymous profile, preserving our identity.

A final warning, even at the risk of being too repetitive: using this type of techniques well to "get into the skin" of another person requires **internalizing** that you are that other person to avoid "leaks". This is often favored **if you have some experience in acting and theater plays**, or if you have previously done exercises or tests of this type that give you enough experience for it. **Do not use this knowledge to attack or harass another person**: your anonymity can be revealed with a court order that aims to a precise point: remember that communication logs are kept, and there are tools and people that can identify you under court order. You just make it harder, **but not impossible**.

AND YOU'RE NOT A CRIMINAL. So do not behave like one. But now you know what kind of resources a prepared criminal can have, so use this knowledge to not to fall into his trap. Someday you can have a position of responsibility and manage money, and that day **you can easily be the target of one of them**. Again: beware.

What if I need more than just personal data to create a fake identity?

If you need to add more than personal data **to your false identity,** such as objects, houses, pets, etc., the use of generative AIs is already so widespread that you can generate anything with them and that, being a random "thing", they will not be able to find on the Internet. You don't have to install anything locally if you don't want to (there are already local AIs of all kinds), you **can use this website** that offers a lot of possibilities about the things it generates: https://thisxdoesnotexist.com/

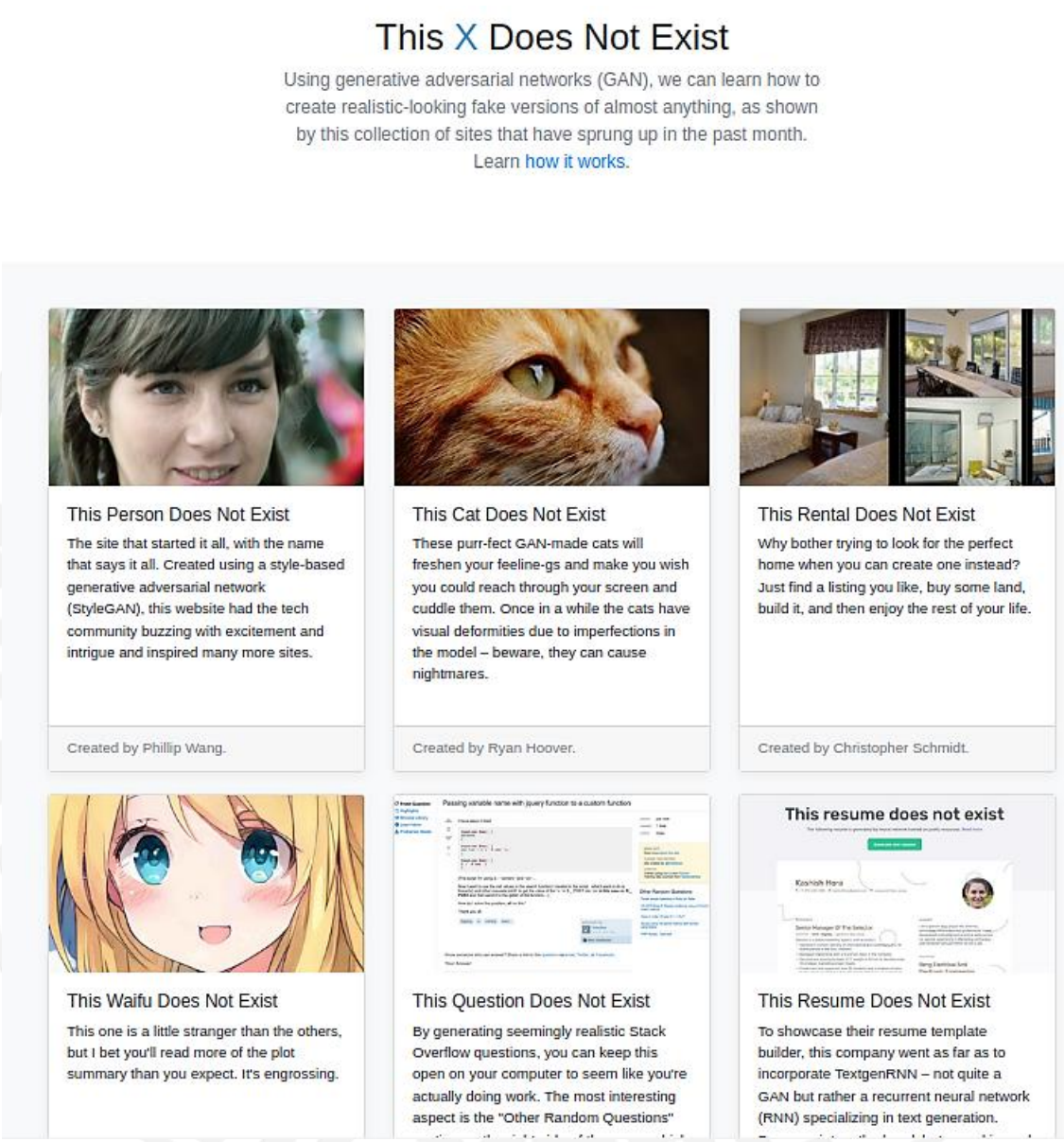


Figure 23. The amount of AI-generable stuff today is overwhelming. And with this website you do not even have to install anything for that...

Another way to generate images adapted to different purposes is ***Canva***: https://www.canva.com/apps/

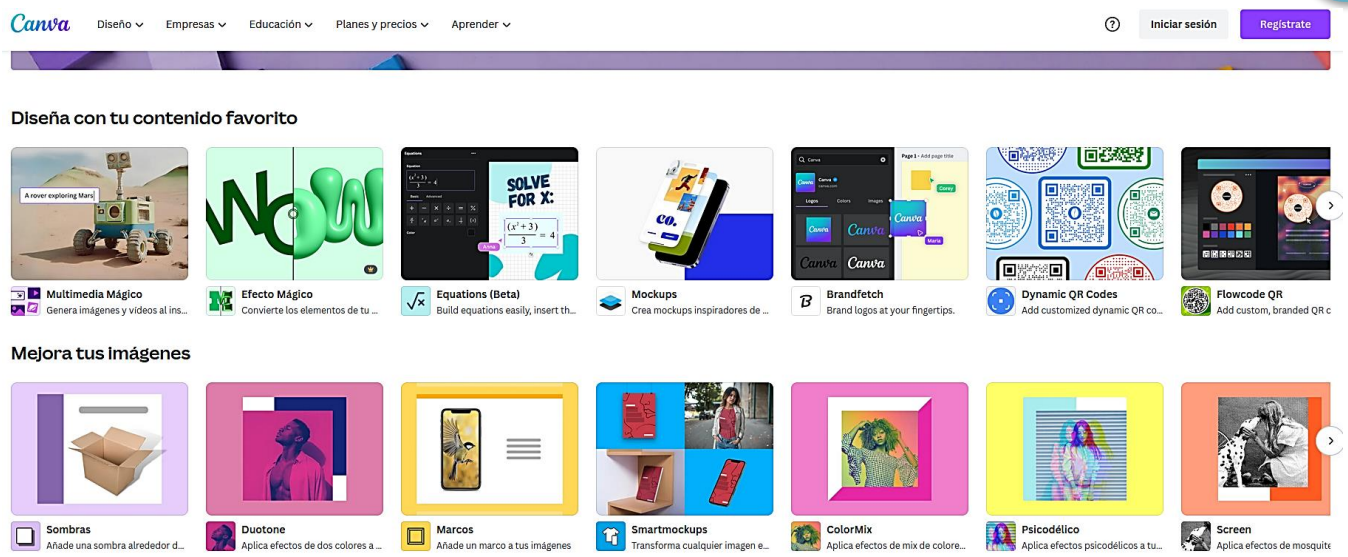


Figure 24. *Canva* is another AI imaging tool that you can use for very specific things you need in research

Machine anonymity

Remote Browsing

This is a technique that **allows you to browse web pages with a browser "from another person",** usually the one of a service you hire that offers you this functionality. Obviously, you maintain anonymity because it will be that other one who "faces targets" for you, and not only on the pages you browse, but also with the possible malware you receive. Therefore, combining a service of this class with a virtual machine provides an **enormous level of isolation** against any threat that comes to us from browsing an untrusted website.

These services are highly specialized and used in highly secure environments, so they **are usually not free.** However, the ***BrowserLing*** service (https://www.browserling.com/) allows us to browse (and also interact with the web pages!) for free for three minutes on a remote machine, with configurable OS / browser, to see possible dangerous websites or in which we do not want to be logged. Simply **remember not to send personal information in these sessions**, as it could be captured and used by the service provider if it has been violated, or as part of their user policies.

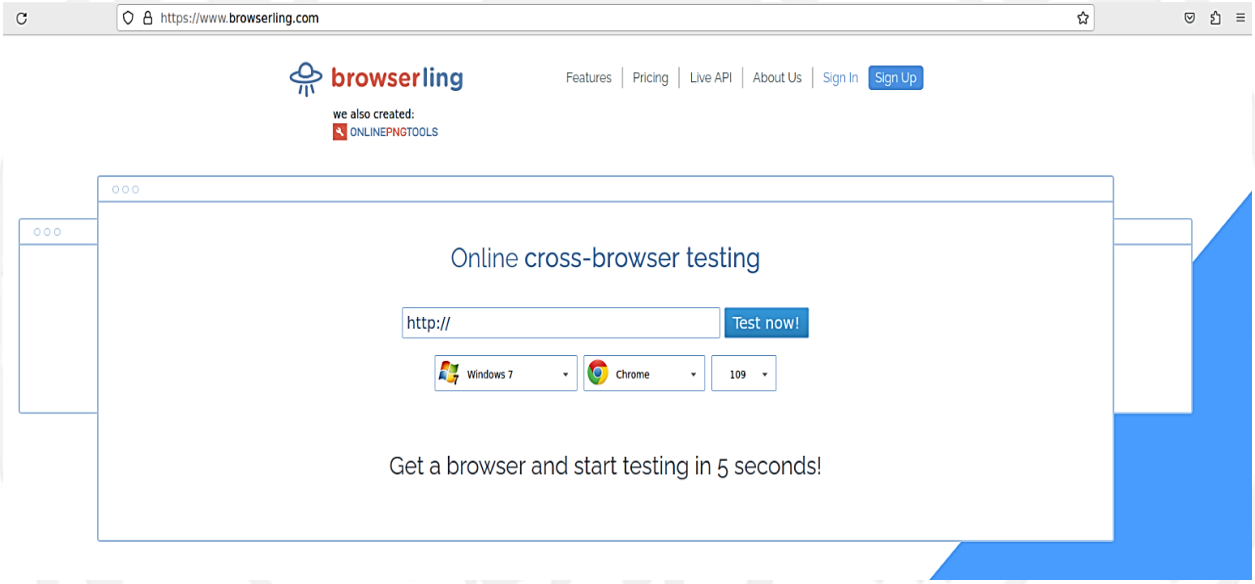


Figure 25. BrowserLing remote browser

Program sandboxing

The execution of certain programs can also leak data about us that they "collect" as part of their user policies or, much worse, **because they are infected with malware of different kinds**. Although running a suspicious program in a virtual machine provides a high level of isolation, it is **not foolproof**, since there are already known cases of malware that recognizes that it is in a VM and uses techniques that take advantage of vulnerabilities in virtualization software to "escape" it. In these cases, for very suspicious executables we can use remote ***sandboxing*** services: Run programs on a remote server that the service we hired provides us and obtain a report of what the program does.

These websites have mainly used for advanced computer forensics and malware analysis and are usually not free. However, there are some that have a free "tier" that allows us to make a limited number of analysis, enough for sporadic usages. One of them is ***Joe's Sandbox*** (seen in seminars), but another is ***Any.run*** (https://any.run/).

Therefore, a service of this class can determine if any of the programs we have to use do something dangerous, because they analyze their behavior and create a report about this, which can also tell us if they "collect" private information that we do not want to reveal.

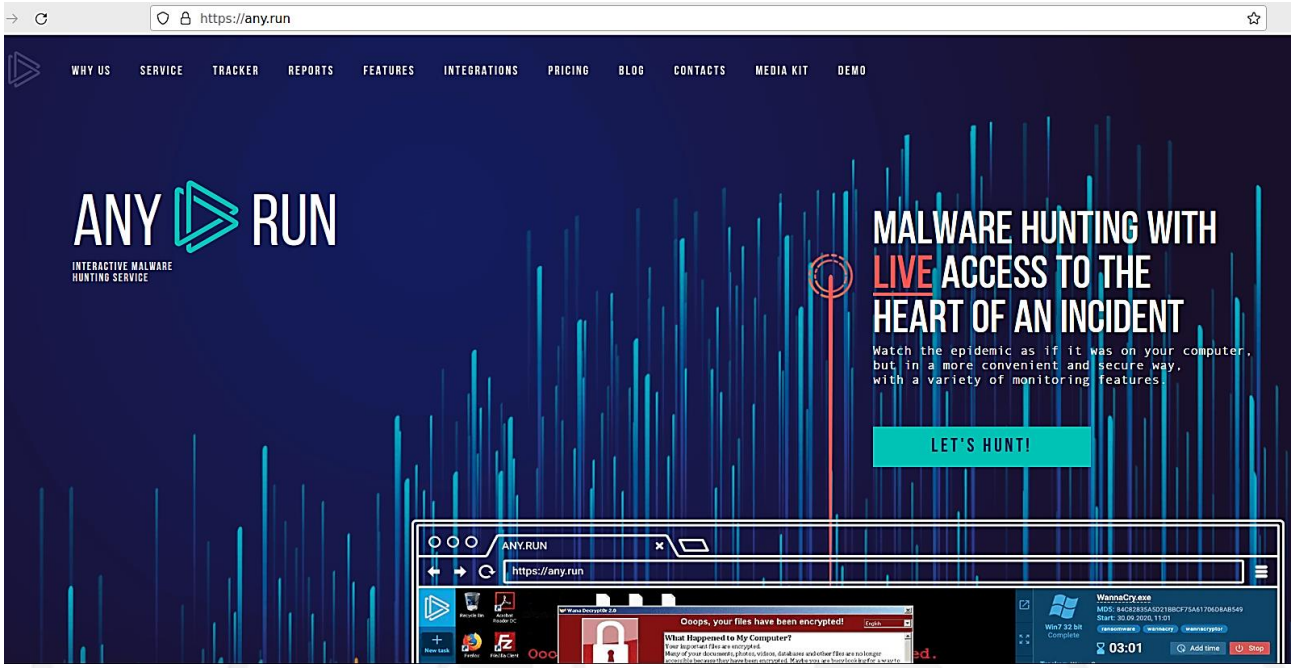


Figure 26. *Any.run* program sandboxing service

VPNs

It is possible to achieve anonymity thanks to a **commercial VPN** service that you can use as a "gateway" between your machine and a remote machine in a country of your choice (depending on those offered by the service you have contracted), so that **it is that remote machine of the chosen country the one that "poses" as you on the Internet**. In other words, in the eyes of those who receive your requests, if the service you hire is well configured, you will be that remote machine that is offered by the VPN service. In addition, being a VPN the traffic between your machine and that server will be private and no one else will be able to see it.

There are many free and commercial VPN services, and some of the free ones are not without controversy because there are suspicions that they sell their browsing data, or the information is not as private as it should be. Of the free ones, one that has a very good reputation is ***ProtonVPN*** (https://protonvpn.com/es/), which although the free version limits the number of servers available, is enough to achieve the privacy we have mentioned (seen on **Lab 7**). It so happens that it is a service that accompanies other interesting ones such as ***ProtonDrive*** (encrypted cloud storage, free up to 1Gb) and ***ProtonMail*** (*webmail* that guarantees the privacy and secrecy of communications between *ProtonMail* clients and, if both interlocutors know a common symmetric key, also with other mail clients).

Figure 27. ProtonVPN

Connections via ToR

The famous ToR network (**Theory Topic 2**) is a free network that **allows you to be (at least theoretically) anonymous on the Internet**. Instead of exposing your IP to the servers you connect to, you actually connect to a network of globally interconnected servers. These servers route each message between them **according to a random communication pattern that changes every X time**, so it makes plotting back the source IP not possible under normal conditions (unless a detailed log of connections is saved, something that "in theory" is not done ... unless there are many nodes "controlled" by entities/governments that do that...). Also, it deletes any information your device provides that allows to identify it.

On the other hand, the traffic between the ToR nodes that make up your "transmission circuit" **is encrypted** between them, so **no intermediate node, only the outgoing one**, will be able to see the actual data you are sending (See: https://en.wikipedia.org/wiki/Tor_(network)). The exit node sends the information to the destination the same way you sent it, so it can see its contents. That is why it is a good idea to also use application-level encryption, such as TLS, to keep your information completely secure.

In other words**, you are deceiving the destination of your communication** about who you are and where you connect from, and all the components of the mechanism you use for it cannot, in theory, know what you are sending. The diagram below describes this in more detail (Source: https://www.f5.com/labs/articles/threat-intelligence/snooping-on-tor-from-your-load-balancer)

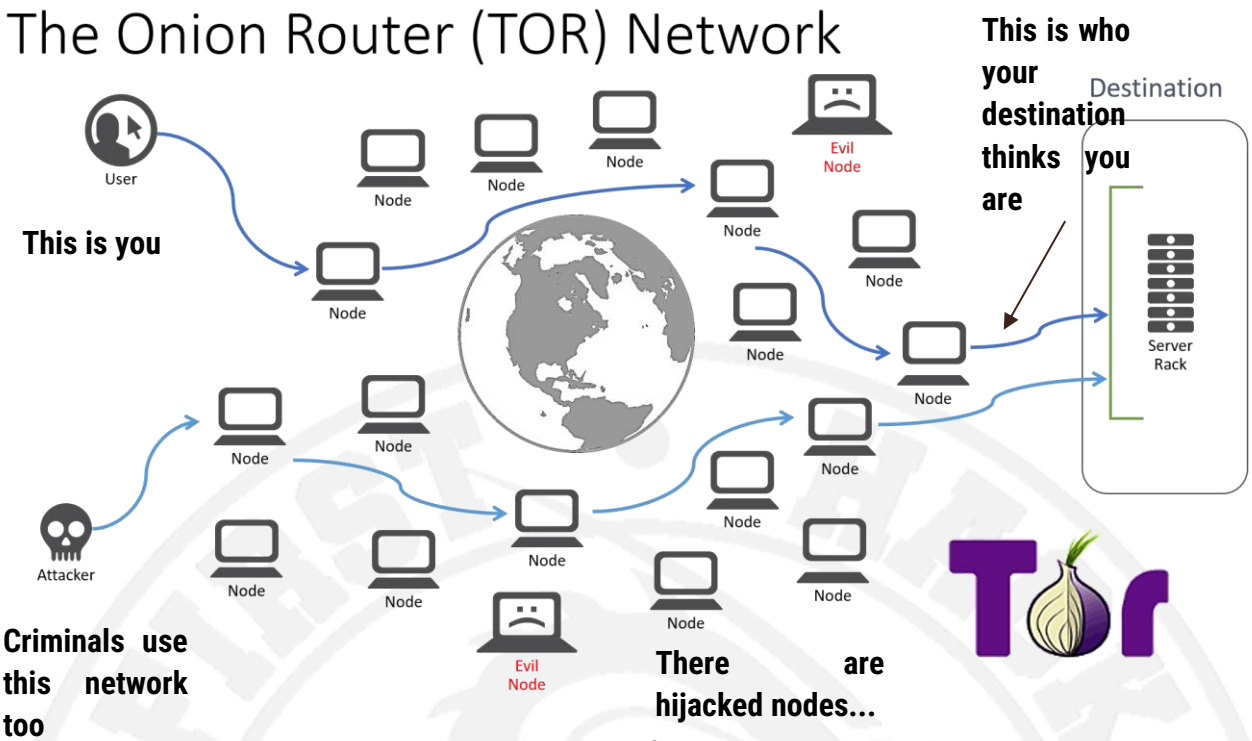


Figure 28. General outline of what the ToR network is

As seen in the image, the anonymity provided by the ToR network can be used both **legitimately** (threatened people / researchers with active techniques who must maintain their anonymity on the Internet ...) and by criminals. In addition, as we said, its anonymity can also be potentially compromised by the presence of nodes intervened by authorities, or trojanized to reveal the data of connections that pass through it.

In **Lab 7** we already saw a way to connect to the ToR network, through the ***TOR Browser***. The problem is that this only anonymizes our browsing and does not do so with the rest of the tools that can connect to the Internet. Now let's fix this, since **the purpose of this section is multiple**

1. Learn how to install the software that allows you to access the ToR network
2. Be able to use ToR with any command that requires Internet access and/or with a browser
3. Check if it really slows down Internet connections to the point of making them unusable
4. Check if there are really known websites that block connections coming from the ToR network.

To do this, follow this procedure:

- You need the latest stable version of Ubuntu right now (or another supported OS of your choice). The tests have been done successfully with an Ubuntu 22.04 LTS. **The latest version of ToR is not working properly at the moment on Ubuntu 18.04**.
- Install the ToR network of the official Ubuntu packages: `sudo apt install tor`
- By default, ToR runs on port `9050`, so check that the firewall (e.g., `ufw`) is not blocking it. On the other hand, you can check if there is an active connection to that network on that port with the `ss` command (I agree with you if you are thinking that the command name is not very fortunate 😊).
- To check that ToR works, the first thing we must know is that **it is a network that is used at will** (that is, on demand, only when we activate its usage). Therefore, even if it is installed, by default every command connects to the Internet as it normally does and, for example, the command `wget -qO - https://api.ipify.org; echo` returns our current real IP.

- If we repeat the same command by prepending `torsocks` (`torsocks wget -qO - https://api.ipify.org; echo`), we should now see another IP, corresponding to an Internet exit node of the ToR network.
- As executing each command prepending `torsocks` is uncomfortable and prone to errors (we can forget to put, and get "discovered" 😟), we can **activate the ToR network for any command that uses Internet** from a given moment onwards by executing `source torsocks on`. If everything is correct, the above command should return the IP of the ToR network, not the real one, even if we do not use `torsocks`. Obviously, `source torsocks off` disables it.
- By default, the activation of the ToR network as indicated in the previous point lasts a terminal session, so new terminals or restart the PC disable its usage. It can be permanently activated with `echo ". torsocks on" >> ~/.bashrc`. Think before doing this: using the ToR network has some disadvantages that we will see later that may not come in handy if you use the machine for other things. If it is a virtual machine that you use just to navigate using ToR, the situation is different, of course 😊.
- Finally, if we change the network configuration of any browser to use as SOCKS Host v5 our own machine (`localhost`) and the ToR port (`9050`), **all our navigation will be done through the ToR network**.

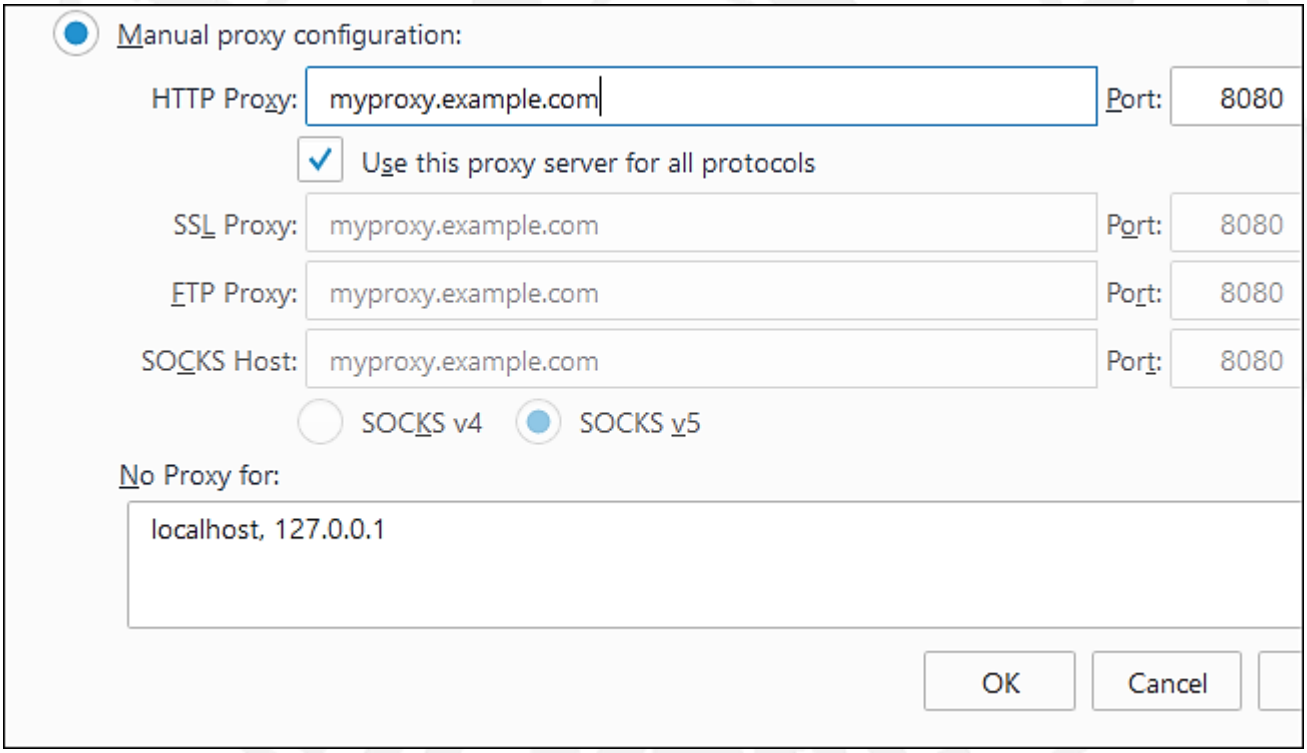


Figure 29. Proxy settings in Firefox

- To verify that you are really connected to the ToR network while browsing you can ask an online service as a https://ipchicken.com/ to identify the origin of your IP. **It should mention a node in the ToR network**.

And what happens then if you "go to Internet" through an exit node of the ToR network? Well, the destination will see an IP that connects to it always, **but instead of your IP it will be one of the exit node** (the one that it is currently serving you), as indicated in the previous image. Plotting your real IP "backwards" is impossible under theoretical conditions, since every X time the machines that make "the hops" that route your connection will change, and they should not keep a log of what was done before. However, the use of the ToR network has a series of considerations that you must consider:

- **Another element of your connection may be revealing data about you**. Imagine for example that you use a website that passes in `POST` or `GET` parameters some data that can identify you as the "source IP" or simply your username. No matter how much "jumps" the ToR network gives, and no matter how encrypted the information between nodes is, it will give your identity away. In reality, any personal information that is disclosed at the application level, HTTP headers, etc. can be used to identify us even if we use ToR, since this network cannot delete it.

  If you need something like this, you should install an anonymous proxy like `privoxy` (https://en.wikipedia.org/wiki/Privoxy, for any app that uses SOCKS) or use ***ToR Browser*** (https://en.wikipedia.org/wiki/Tor_(network)#Tor_Browser), if you're browsing. It is very easy to install it on a standard Kali Linux, for example: https://www.kali.org/docs/tools/tor/ . This browser already uses ToR by default when you launch it, it is not necessary to configure it as a "standard" one. It is also a bit absurd to use a service where you have to identify yourself as your real user through a ToR network, unless what you have to hide, more than your identity, is your location. Of course, your identity may be false like the ones we saw in the previous section. In that case this does not matter 😊.
- **The connection speed will decrease**: This is logical if you think about what the ToR network does, which makes your data "jump" multiple times between nodes. Each "jump" is an extra delay, but it is a price to pay, and for "normal" navigation it is not something limiting.
- **Some pages will not allow you to access them via ToR**: Yes, since the ToR exit nodes are known, some websites (for example banks, large companies ...) maintain an updated list of existing exit nodes at any time and block any access that comes from these nodes or lacks certain values in the HTTP headers that may have been deleted, **assuming that any connection coming from such sources is inherently malicious.**

  Against this we cannot really do more than try to make one more "jump" between the exit node and an "intermediate" machine at the destination, but obviously it cannot be associated with us because otherwise all this is meaningless. I guess that now you can imagine **why stealing any PC from a home is interesting for many attackers, right?** 😊. Obviously, in this course we are not going to reach this level of sophistication, renting with an anonymous prepaid card and a false identity (as in the previous section) a virtual machine in a cloud provider to make that last "jump" (***TA0042. Resource Development*** of the *MITRE ATT&CK*, **Topic 7**).

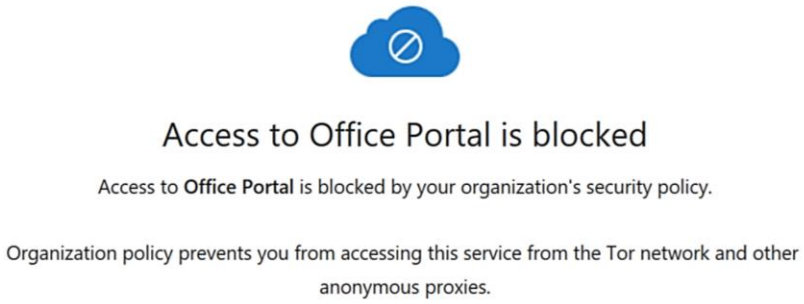


Figure 30. This company has decided that nothing belonging to its Office 365 cloud infrastructure will be accessible via ToR or proxies that anonymize the connection

Searches from another location

If what you need is simply to search from another location and you do not care much about the privacy of what you are looking for because you only need the results, all this is too complex just for that. You can instead use ***I Search From*** (https://isearchfrom.com/) to search by pretending to be in other locations or using different devices. Sometimes you find information just like that (like pages with a mobile version that give more information than their PC equivalent).

Figure 31. If you are just interested in pretending to search from another site or being a different device, this tool works

🕵️ Preparing the OSINT study. *What kind of information should we look for?*

Once we have introduced these techniques, it is clear that thanks to them we will be able to access many, many data from a large number of people, entities, or companies. Obviously, accessing all this data may initially seem a very complex task (because it really is 😊), and to facilitate this work it is necessary **to give order to the information that could be obtained from a person or entity**. In other words, organize yourself and have a diagram that guides you about **what things are going to be interesting in your audit scenario**, instead of "going crazy" and obtain information that may not serve you at all, or has no application in your specific case. A possible set of information is represented in this image:

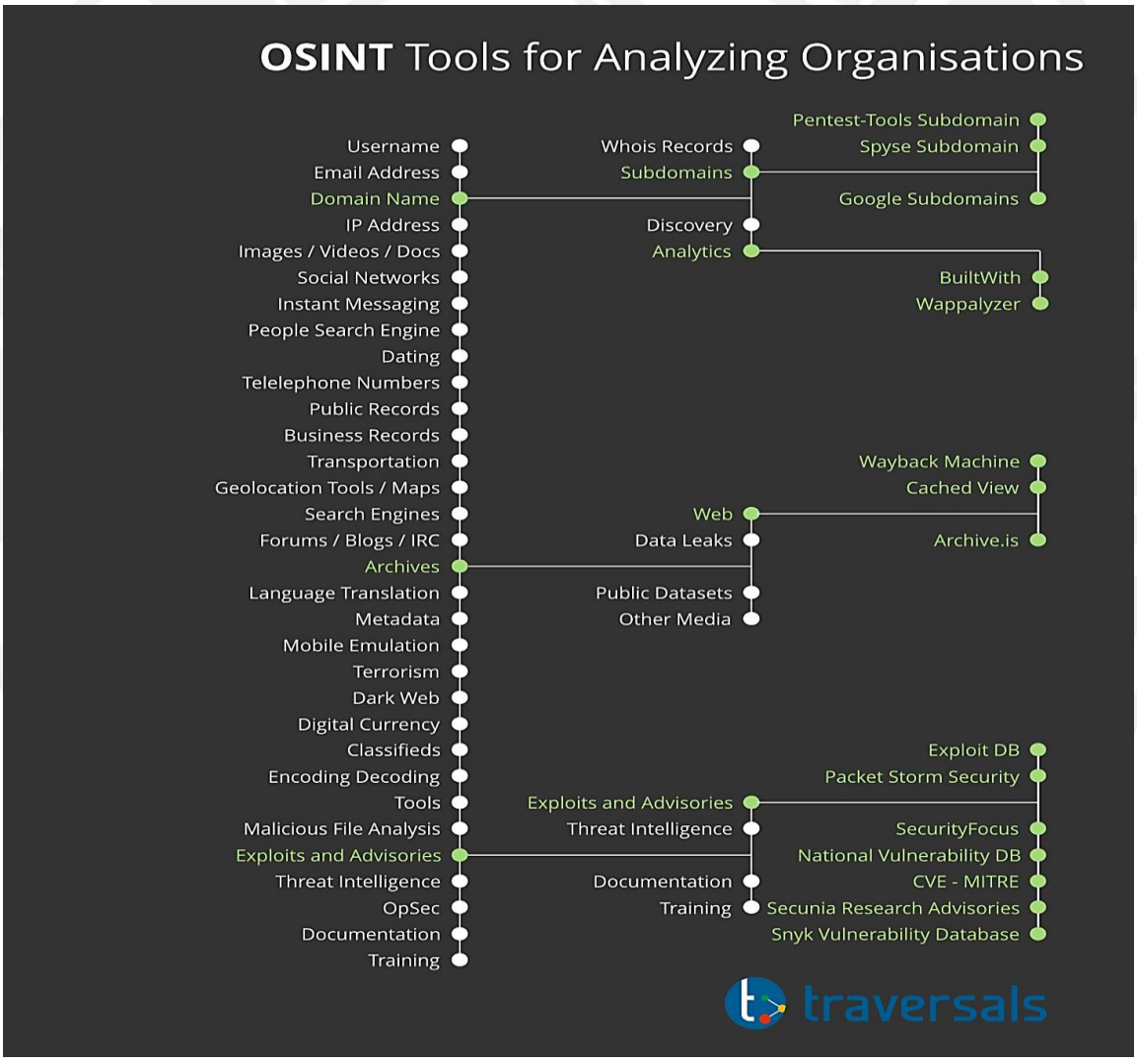


Figure 32. Example of personal information that can be located via OSINT. Source: https://traversals.com/blog/osint-tools/

But we say possible because **there really is no "canonical" order**, but different ones can be used depending on the type of objective or the country in which you reside. For example, this is suitable **for a US citizen** since it mentions data that only makes sense or is publicly available there (although you will agree with me that much of it can be applied to a Spanish one).

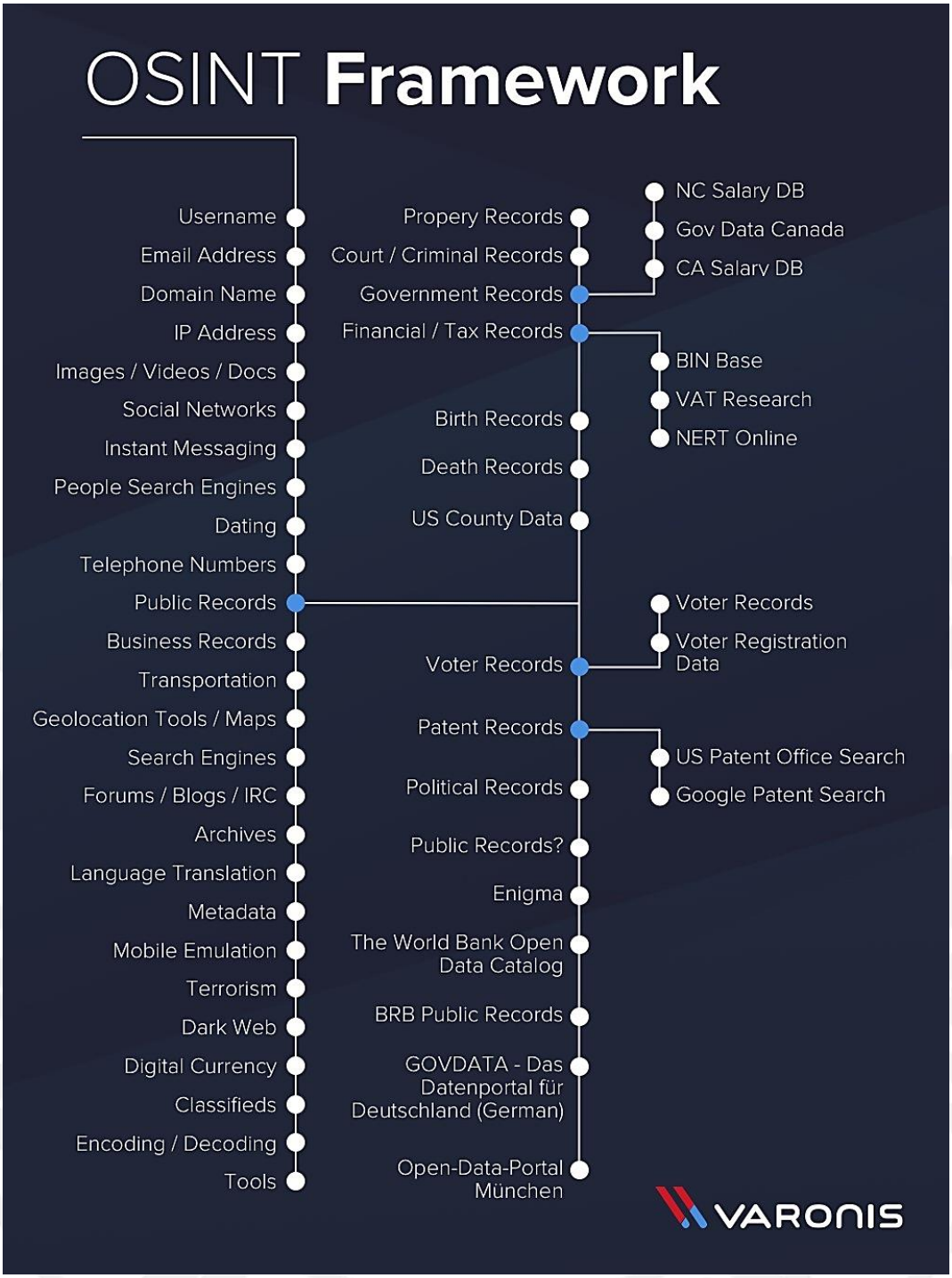


Figure 33. Example of data obtainable from a U.S. citizen: Source: https://www.varonis.com/blog/what-is-osint

If the audit is **company- or organization-oriented** and more general, this abbreviated scheme may turn out to be more appropriate.

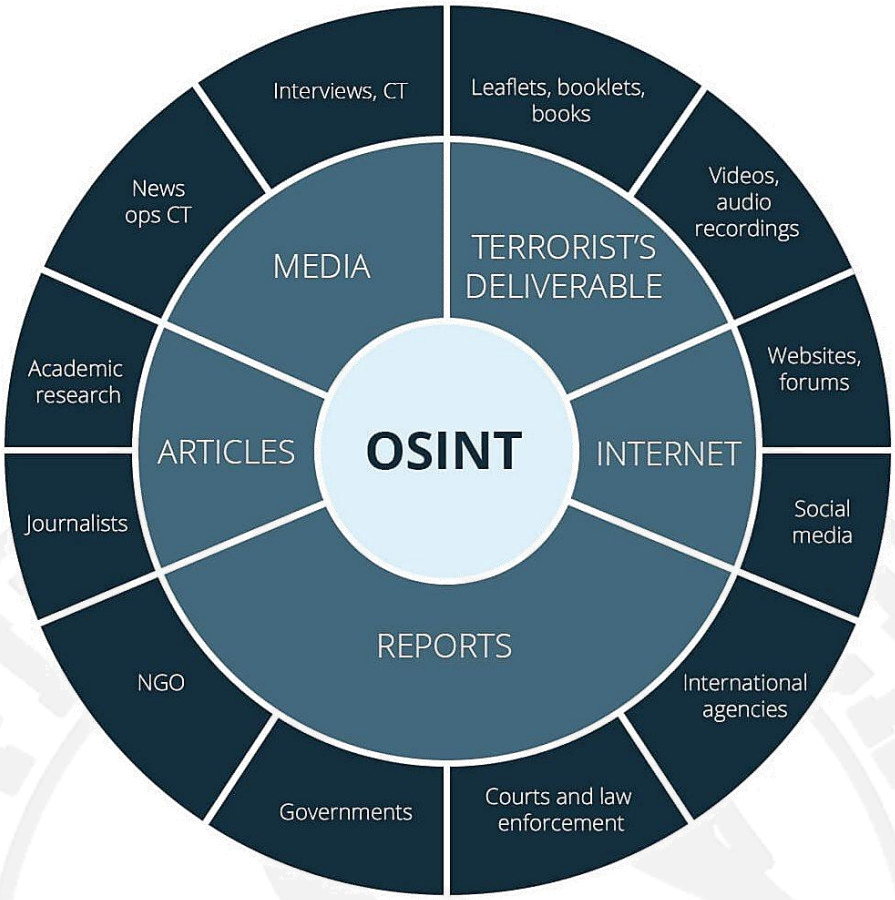


Figure 34. Abbreviated scheme of OSINT information to be obtained from a company or organization. Source: https://www.defensa.com/ciberseguridad/osint-como-obtener-inteligencia-explotando-fuentes-informacion

And if we want an **in-depth study of technical and "human" aspects at the same time** (and the time to do the research allows us), we can use this *PortSwigger* scheme (https://portswigger.net/daily-swig/osint-what-is-open-source-intelligence-and-how-is-it-used). We will base the rest of the document on this one.

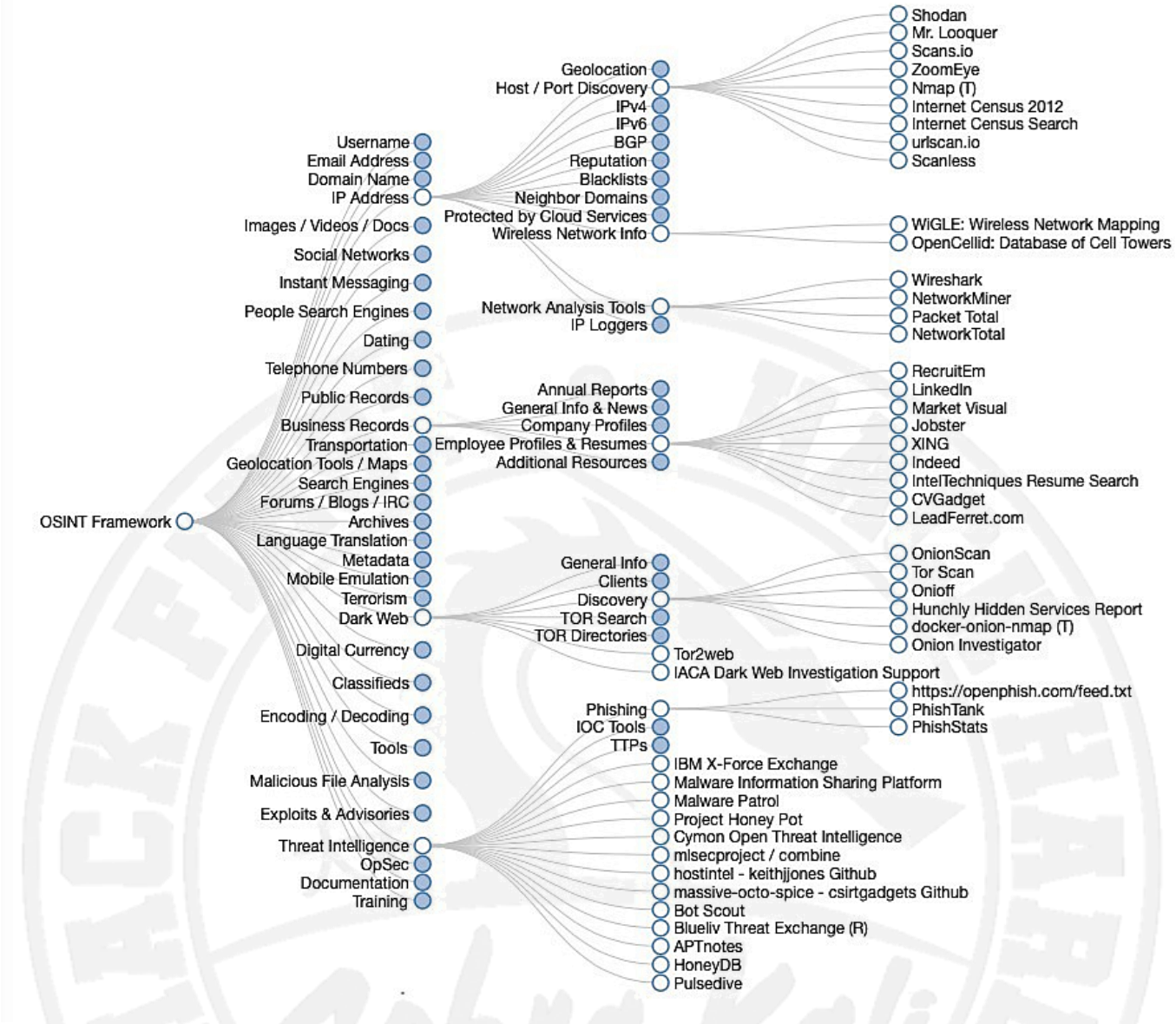


Figure 35. *PortSwigger* schema of technical and personal information obtainable on an OSINT investigation

On the other hand, your interest may be to find out only **technical data of a specific business and the personnel associated with it**. In that case, this diagram can be useful to guide your research:

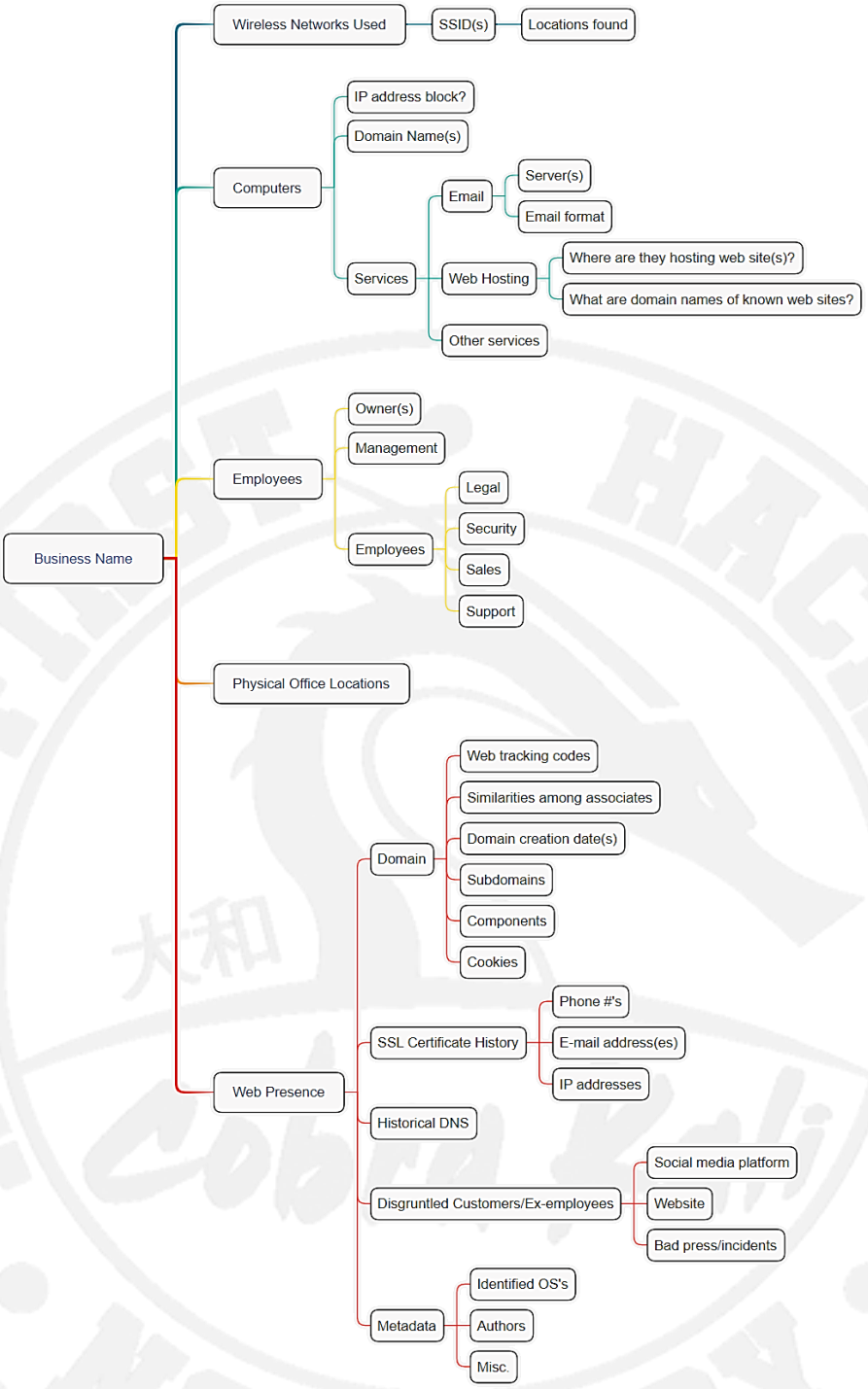


Figure 36. OSINT data associated to a business. Source: https://es.linkedin.com/in/asier-rohde

As you can see, all this is about:

- **Obtain a certain set of public data** (more or less extensive, according to your needs, and always in a previously planned manner) of a person, entity, organization and / or their relatives.
- Once this is done, **it is classified** by entity, person, or type of data
- Once you have all the information, **you analyze and draw conclusions from it**. From these conclusions one can draw everything: possible valid usernames in a system, indications of passwords, aspects of personal and work life not obvious a priori...
- This is often used for phishing tasks of different kinds or to facilitate initial access mechanisms (***TA0001. Initial Access*** of the *MITRE ATT&CK*, as we see in Theory **Topic 7** and in **laboratories 11**

and 12), as it is one more reconnaissance technique. Although, as you can see, depending on how much you "stretch" it can encompass many others....

And how do I get all this information? Well, the truth is that **there is no automated method that works in all cases**: it depends on the victim, their context, and their environment. What is certain is that in the vast majority of cases it is done **with a combination of manual and automated searches**. Moreover, manual searches usually "feed" automated searches, although both must really act as a complement (a manual search can serve to confirm that an automated search has found something real and not a false positive).

Next we will see manual techniques to get information and their possible practical application, relating them when necessary with the rest of the course. Once this is done, we will see a reference tool to collect that information and automate searches: ***Maltego***. Finally, we will see techniques to get this information "by any means", since sometimes the person who hires us gives us additional possibilities to extract information, such as doing **simulated phishing campaigns** against a target to check their "cyber-awareness" (I am not kidding, many companies do them, Uniovi did one a few years ago). With all these techniques, your "cibermarujeo" ability will be increased to more than 9000 points 😉.

"Manual" OSINT: The art of searching and interpreting data

OSINT on employees or individuals

The amount of information that can be found about a person on the web is often **overwhelming**, and therefore if we do not know where to start we can feel lost. In this section we are going to talk about many different kinds of information, but a **general outline** can be found in this diagram, to give you an idea of how exposed we can be on the Internet.

Figure 37. We are not aware of the amount of information we expose on Internet...until we look for it

If this seems like too much for you, **another more concrete scheme** can be this:

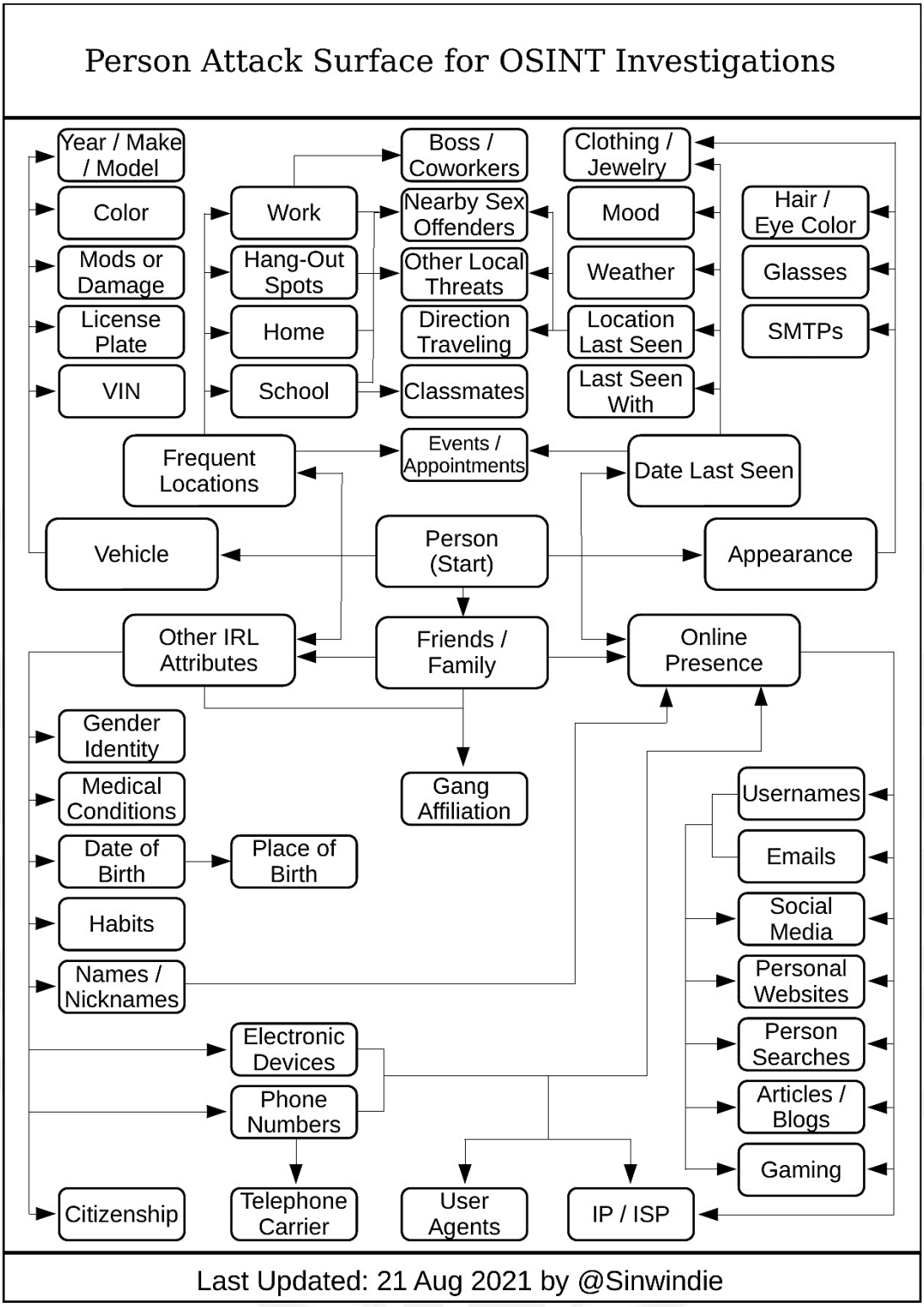
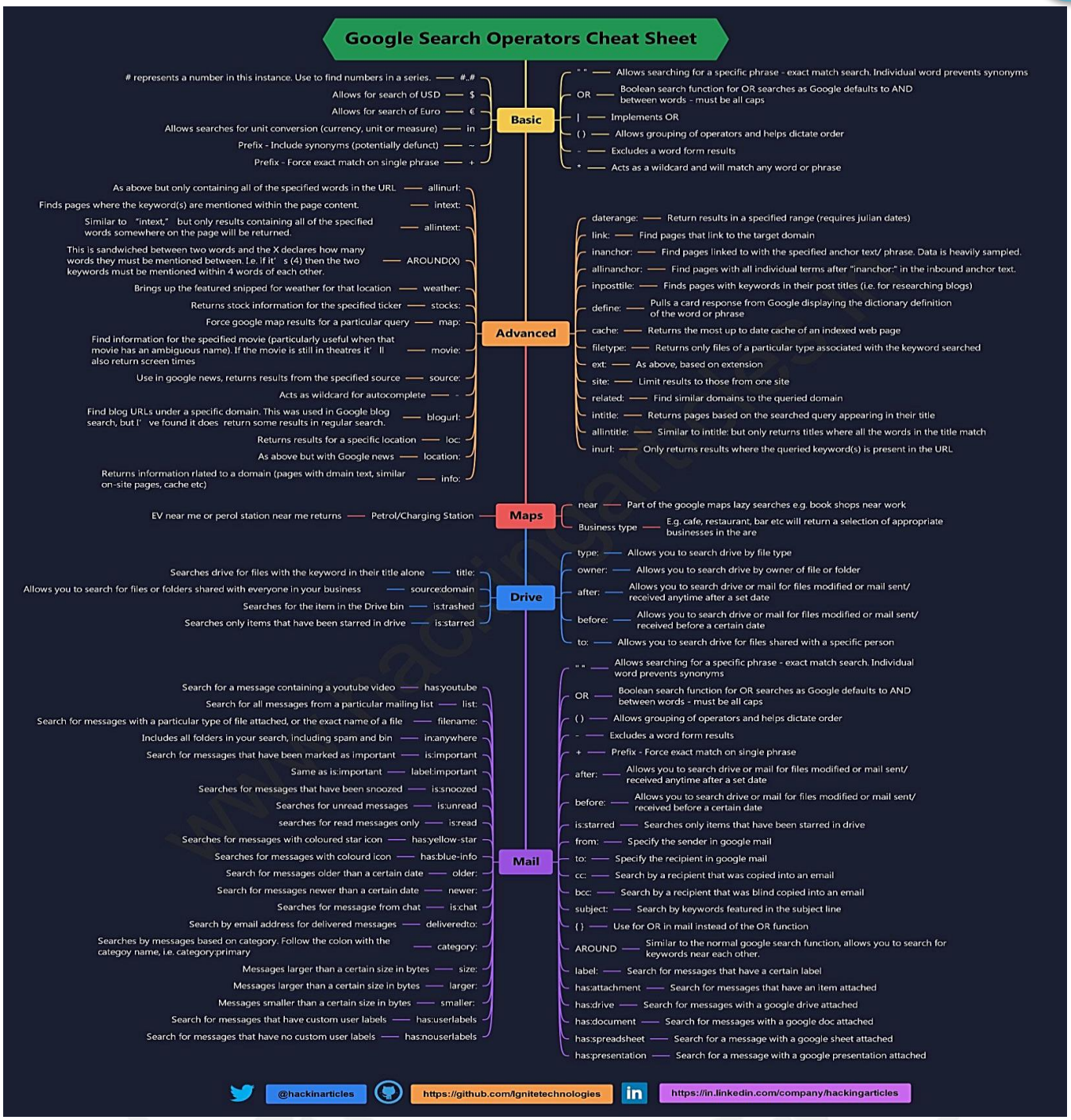


Figure 38. More concrete information about what can be known about a person on Internet

Things about persons in general: Names / Images / Videos / Photos...

The first thing we need to know is that **we should never underestimate the power of a search engine**. It is not about searching for a name and that is it, *Google*, for example, has many search options:



Searching for someone's name online is a basic operation, but we can improve these searches with simple tricks:

- Search for the full name if available.
- Enclose the full name in quotation marks to find exactly that name. Do it with the format `<First Name> <Last Name>` and `<Last Name>, <First Name>`
- Add profession, location, hobby, ... or any other relevant data you know after the name to improve the accuracy of the search.
- Use ***Google Advanced Search*** (https://www.google.com/advanced_search) to include or exclude certain words, phrases, or websites from search results.
- Search for the name on social networks, especially on *LinkedIn* because it is the one that can give more potential data.

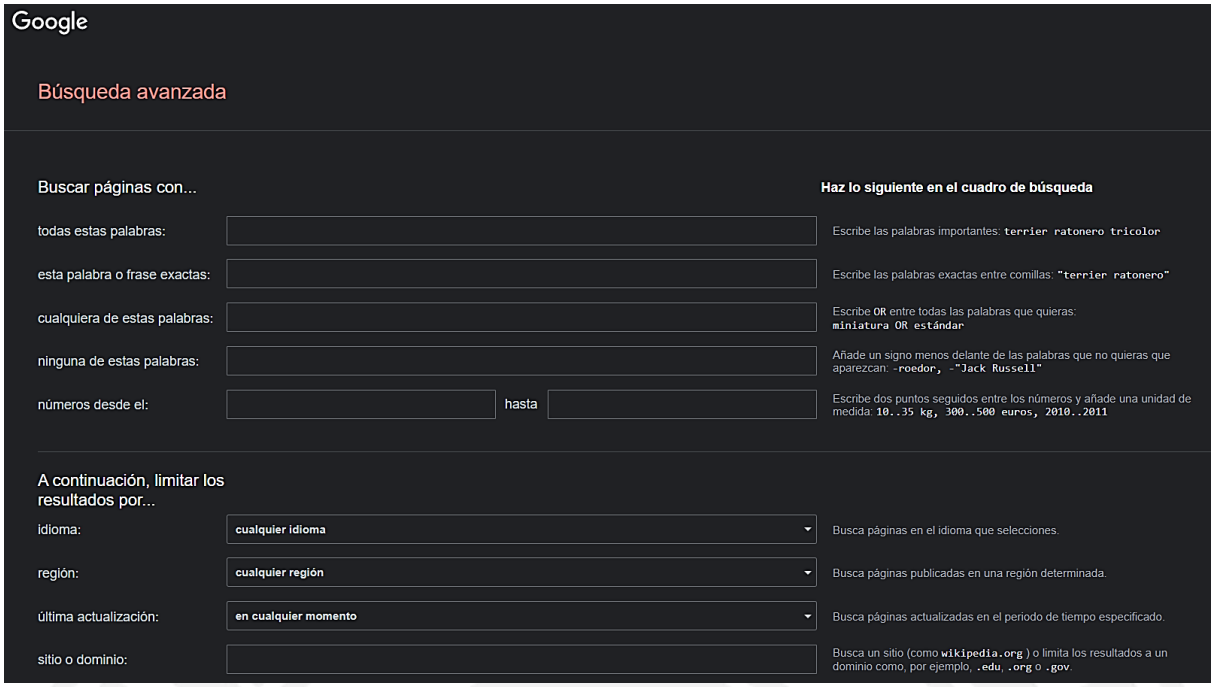


Figure 39. *Google Advanced Search*

If we don't find the information we are looking for right now, we can also use ***Google Alerts*** (https://www.google.com/alerts?hl=es) to be notified when *Google* indexes certain search terms we can configure and that may be interesting to us, so we can know when certain data appear on Internet. This is also advisable to know when **we** appear on *Google* search data, to prevent data leaks, unwanted mentions, or similar situations.

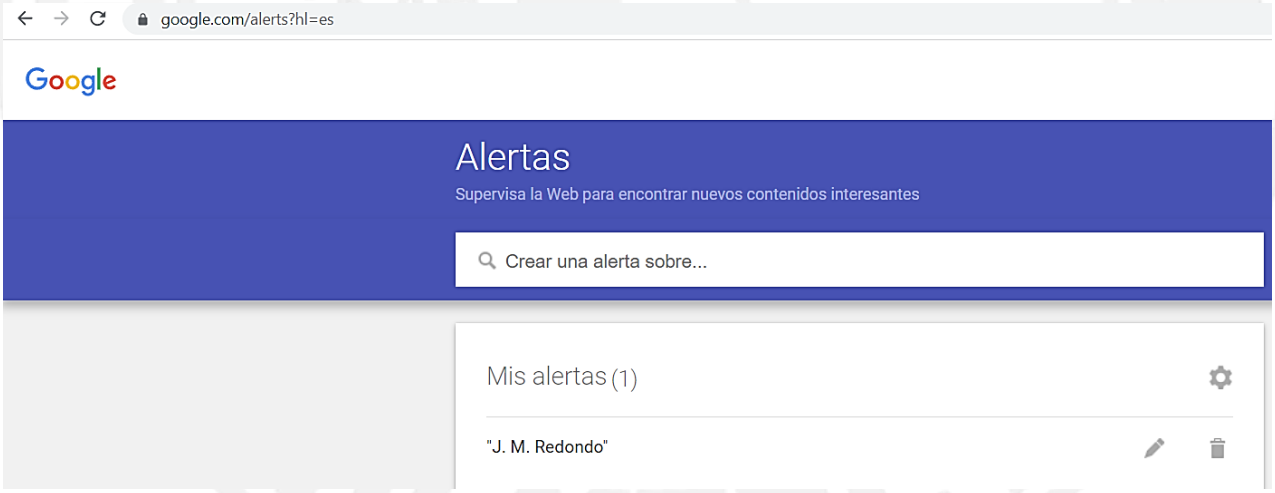


Figure 40. Google Alerts service

If we do not have any data of a target beyond a photo, the best way to find more information is to resort to a reverse image search with *Google* (https://www.google.com/imghp). Simply download or crop the image you are interested in (Windows "***Snips***" app, for example), save and drag and drop it to the search box.

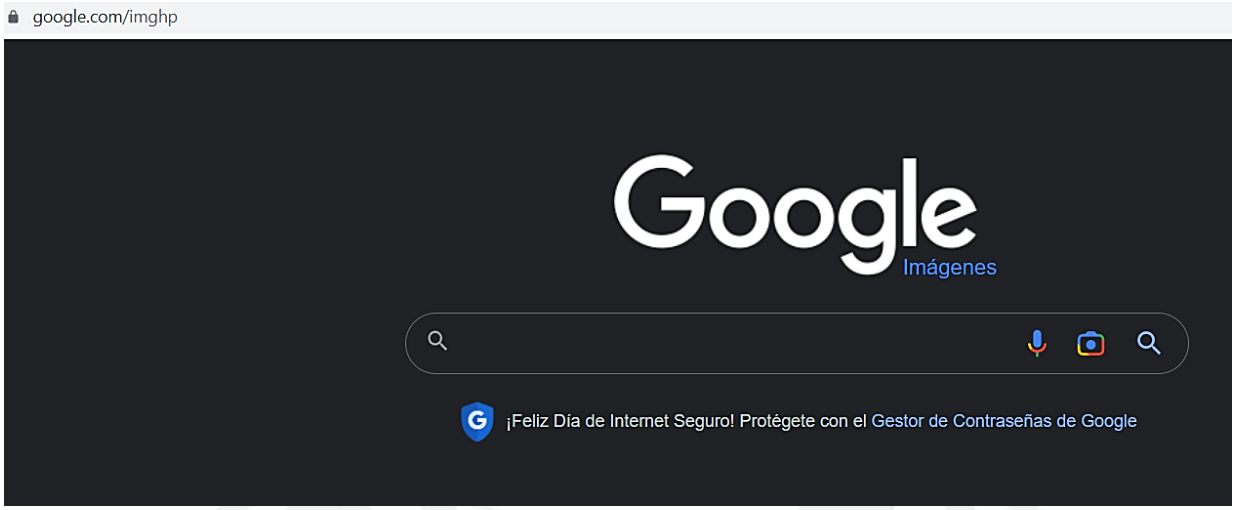


Figure 41. Google Image Search

With a little luck we can have a name and the combination of this with the image probably allows us to find social networks of the objective. With this, it is common that more interesting data will appear, such as names of people associated with the investigated one (relatives, co-workers, partner(s), descendants, etc.) to "pull" their information and continue with the investigation. Every person found like this is searchable to know what we can find out from it. Other OSINT tools can do **facial recognition** via photos on social networks, but you must locate one that is still maintained in the ***CiberPatrulla*** repository.

A photo is something that can offer us much more information:

- For example, unless it is uploaded to a typical social network (since they delete them) it is possible to examine its **metadata** for more information (we saw it in **Lab 2**, *remember?*). Although social networks eliminate metadata, it is also possible that the same photo present in a social network of the target has been uploaded to different sites (that is, you have to use *Google* image search again). *Why is that interesting?* Well, because some of those sites may not remove the metadata, and if the target does not do it on its own, there they will be. Therefore, there can be the same photo with and without metadata on the Internet, depending on where it is being displayed. Thus, it can be the paradoxical case that you can know a lot about the metadata of a photo in an RRSS by a copy of it outside that RRSS.
- Even if a recent photo does not offer metadata, perhaps **a photo from the past does** (the concern about deleting metadata is something that became popular recently). *What do you know that can search through the past of a website on the Internet and, within it, in its individual archives?* Exactly, the ***Wayback Machine*** website of **Laboratory 2** 😊 .
- On the other hand, **details** of the photo such as **recognizable places, date of the photo** (sometimes it is printed on the photo itself!), its name (but beware, many sites such as social networks change their name for something random), car **license plates**, bus / train / taxi stops **numbers**, specific **streets or squares**, **monuments** and a long etc. can help us, with a skillful use of *Google Maps/Earth*) to geolocate a person at a given moment in time (or approximate) or any other relevant information.

  This is reasonable, if you want an example you can look why a well-known streamer was in sock (*wink* *wink*) when the address of his new home was known by a detail that appeared in a window in the background of a photo that he uploaded to his RRSS. Sometimes researchers use an anonymous identity (like the ones created in the previous section) to **ask people** (also anonymous) **on different forums** if they know where a photo belongs. The odds of finding someone who lives nearby or passes by are high, especially on large forums. On the other hand, imagine that a photo has the camera model with which it was made, and that it is a very expensive professional camera.

How many people can have that camera among a list of possible "suspects"? And if it is a very expensive camera, it is very likely that it "boast" about having it in RRSS ... *do you understand a little how things are going?* Yes, if it looks like a detective game it is because, sometimes, it is a detective "game" 😊.

And why do you want to know that? Because the more information you have, the better your research will be. A person who likes to "photograph" his life more or less intensely is exposing, perhaps unknowingly, a huge portion of his life. If you want real examples, I can tell you about a few 😊

- If you usually take pictures **in different places and times with high frequency**, it is technically feasible to make a "map of your routine" with *Google Maps/Earth*. Apart from the risks that can have for people to have places where it is more possible to find them at a given time (creeps, stalkers ...), including doing "live espionage" if the objective is important enough, establishing a routine on a map can also reveal workplaces, recreation places, gyms, or even one's home, even without having geolocation activated. Obviously with geolocation activated this is much more trivial 😊
- If the person is used to putting **vacation photos**, skiing in familiar places, mountain, etc., he is not only revealing his tastes, but also where he usually goes when long vacation periods arrive, again increasing the options of finding him/her in a place in the real world on certain dates ... or where not to find them.
- These types of social photos **can involve more people**. Each of them is an additional objective to investigate following the same techniques that are used with the "primary person". *And why do you want to do that?* Because frequently these close people write things related to the first, their activities, or their work that are useful for research. *Do you want to avoid revealing you are going to get married?* Bad luck, your friend has taken a picture of you in the background at the banquet and uploaded it to *Instagram*. *Do you want to keep your children private?* Sorry, your son's best friend's mother took a picture of them in the park and put it on *Twitter*... etc. The worst thing is that the "victims" of this type of problems have a very difficult time avoiding them or defend themselves from them. At the social level there is very little awareness of the dangers of this type of activity.

  Keep in mind that even work anecdotes can be used for a **spear phishing** attack, for example. Of course, this allows us to build a network of relatives as we said before, which is still another powerful source of information. That network can be built and enriched with the *Maltego* program that we will see later.
- A photo can also take **physical characteristics**. For example, posting a full-length photo with something next to it of recognizable size allows you to calculate the height with a simple rule of 3, such as the man holding an A4 sheet of the next photo, which measures around 1.77m. There are even AI tools from these photos that allow you to calculate the approximate weight (in this case between 84-87Kg).

Figure 42. Finding out the height of a person with a photo. Source: https://ciso2ciso.com/introduction-to-doxing-osint-methods-for-information-gathering-by-hadess/

And now you may wonder. "*Okay, very good. Now tell me what I can do with this type of information*" Well, look, it is a matter of letting your imagination run wild... I tell you how I let mine fly, which can serve as a clue 😉

- If the **height varies slightly over time**, you may notice that the person has a back problem, such as spinal compression or low back pain that forces you to walk slightly hunched over.
- If the person is **very tall or short**, he may have complex or problems finding clothing sizes, which can serve as data for a phishing operation or know the type of stores he frequents in very specific cases.
- A **drastic weight variation** over a relatively short period of time, both up or down, may be indicative of:
  - A **sedentary job** (new or current)
  - Having **left the gym**
  - Being on **certain medication** where weight gain or loss is a clear side effect
  - An **illness** (usually serious: anorexia, bulimia, depression, and other mental illnesses (medication against them often causes these side effects))
  - A **change of lifestyle for the better or worse** (preparing a demanding race, a health "scare" that forced the person to be put "on shape", a complex stage of life ("40's crisis"), study an opposition, or even a sentimental breakup.

Although they are possibilities to introduce in the "mix" of information about a person, I agree that **they are assumptions and some very risky**. But at least, they are topics to introduce in subsequent conversations with the investigated person that we may have (remember the section of false identities in cases where there is engagement), to use in a phishing campaign later, or to contrast with other information that we obtain, to see if any has a chance of being correct. As I say, the limit is in your imagination 😊 .

Sports of interest

Photos with flags, scarves, gyms, etc. of football/basketball/etc. say a lot about one person. If a person is photographed doing different sports or in various places, not only do you **have a possible place to find them**, but you have one more topic of conversation or tools to use in phishing. Of course, there are also other people to investigate as a result of such photos.

Figure 43. Person who publishes his sports life in RRSS. Source: https://ciso2ciso.com/introduction-to-doxing-osint-methods-for-information-gathering-by-hadess/

This is more important than it really seems. If you are a "true fan", for example, of the Sporting or Oviedo football teams, you take pictures on the field, going to the matches, etc. you are actually saying where you are going to be (and, beware, also **where you are NOT going to be**) with a great probability, in public and known moments of time. If you take pictures following your child's amateur team at their matches, those match schedules are easy to find, so they will know where to find you regularly. The composition of the teams is also usually public (often, even with photos), so they already have your children and their friends. The staff that is part of the team is also easy to obtain, so they already have relatives with whom to find you "jumping" through RRSS profiles if your photo does not tell them anything. And a great etc., if you think about it ...

Cars (or other possessions)

Do you have a cool car? Good for you, seriously. But if you take a picture that shows the license plate it can be used to know who you are and more information about you. Apart from, of course, estimating your purchasing power. This is unless the person is a "poser" who rents them to "brag". In that case the license plate, stickers, or distinctive characteristics of the car (which make it recognizable in photos of other people) can serve to know that the person is precisely that :)

Figure 44. Cool car! This type of car tells me that you have on children... . Source: https://ciso2ciso.com/introduction-to-doxing-osint-methods-for-information-gathering-by-hadess/

Obtaining the data of a car owner by their license plate however is not free and requires providing a justified cause, as indicated in these two links.

- https://tramicar.es/blog/como-identificar-al-propietario-de-un-coche-por-la-matricula
- https://blog.reale.es/averiguar-propietario-coche-por-matricula/

It would not be surprising that one day (if it has not already happened ...) **a database of cars registered in a country is leaked** where you can directly search for your data by license plates (remember: we are actually thinking about how criminals act, if having ethical and moral limits you can already find out a lot, imagine if you do not have them ...). However, many times **it is not even necessary**, because it is the person himself who, based on analyzing all the photos he uploads to the networks, gives this information himself. This is an important aspect too: a single photo could be harmless, but **the combination** of that photo with others can give us interesting facts. And normally those investigated do not consider the context of their past photos before uploading a new one...

Food

If you take a picture of what you eat, it may reveal not only which restaurants you frequent, but you can make assumptions about your country of origin or race, intolerances, or whether or not you are vegan, for example. Apart from other elements (people, objects ...) revealing that may be in the photo, of course 😊

Colors / Clothing Style / Self-Care

Your favorite color can be used in spear phishing activities (there is a whole marketing discipline that uses subtle data from the human mind to reach people ...). In addition, the type of clothing can say a lot about your personality, economic level, or even stores in which you frequently buy (physical or online) if they are of specific brands. Like everything you reveal, you do not know when it is going to be useful data...

If the person wears hairstyles that look like the ones that they are complex to do, change the cut style frequently, wear groomed beards, makeup of a certain style, etc. He is also revealing facts about his tastes, personality, possible job, interests, or intentions, and, perhaps, places he frequents.

Figure 45. Person who likes himself very much. Source: https://ciso2ciso.com/introduction-to-doxing-osint-methods-for-information-gathering-by-hadess/

Books / Movies / Favorite Songs

Publicly revealing your favorite books/movies/songs, including any comments you may make about them or their characters in public forums, can be very revealing of things like:

- Educational background
- Political leanings
- Religious faith
- Vital values or thoughts about issues that may be controversial for certain sectors of the population if you are for or against them.

In other words, **it is very easy for you to "socially catalog" yourself in this way**. While it is true that it is very common for people to give that kind of information "for free" on social networks. For example, there are people who talk about their membership in a particular political party as if it were their family's coat of arms...

Although it may seem silly, revealing *Spotify* lists publicly and repeatedly, the song you are listening to right now (and the comments you make) can be used to infer your mood, upcoming major life events, illnesses you are going through (depression, anxiety...), and personality traits. I am serious. It is not the first time I have seen lists like "`Songs to overcome a breakup`", "`Themes for the wedding`", "`Background music for the to study MIR`" or things like that. Besides, you know that the application itself classifies for you songs based on the mood you put on it, so you can also make those kinds of assumptions. All this can be useful information in spear phishing, for example.

Photos of your house

As it usually happens (and it has been in the news recently with a streamer, as we said before) taking photos of your house can easily reveal **where you live** by the background of these photos (if recognizable), or other details of your social and personal life. A background photo that shows **someone's face**, **pets** (nameplates, names that may be indications of passwords that you use on your websites ...), **type of devices** you have (which can give clues on how to attack them from the outside simply, for example, finding out their brand and model by what appears in the photo), or **any object that appears by accident** in the photo and reveals something about you (it is not the first time that confidential documents, IDs, credit cards... appear), yours, or something you have. For example, a company recently boasted on social networks about its new routers installed in its data center. The photos showed the model and brand clearly, and that model had a recent known vulnerability. *Have they been patched?* I think you understand me...

The most striking case is that of "window background". Think that although the background itself is not easily recognizable, if you add that data to what you have already been able to find out (routes, routines, etc.) you make locating your house much easier. Now add that you are going on vacation on a specific date, and you understand why this is dangerous....

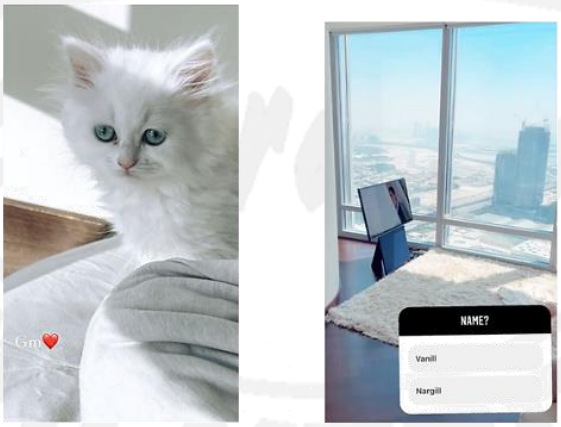


Figure 46. Window background photo that can lead to recognizing where you live. Source: https://ciso2ciso.com/introduction-to-doxing-osint-methods-for-information-gathering-by-hadess/

Photos of your house / other properties keys

Posting a photo of the keys, for example of your new bought or rented house can seem very nice (in fact it is), and it is something that is frequent in social networks. **Unfortunately, it is a very bad idea**. These photos **can be used to make a functional physical copy** of the keys you just photographed, with the security problem for you and yours that this entails. If you also consider that the photos may be tagged with your location, may contain some metadata, or the context of your RRSS can be examined to deduce where you live, you will understand why this can be a serious problem.

A modern 3D printer can copy a key. The reason is that mobile photos today have a very high resolution, and therefore it is feasible to translate them into a printable digital scheme of the keys. If you then use elements of the photo such as hands, rings, or other background elements to get an idea of a scale and dimensions of the key, then you eliminate another problem. Although keys are designed with different lengths and depths of ridges, the shade of these can be used to estimate their size. In addition, if they go against you for some reason, since 3D printing today is very cheap, they can make several versions of the key until you find the one with the right depth and length.

Many keys also have **a written code**, which can be used to obtain an exact copy of it from the manufacturer. If someone is able to see such code by a photo, they can use it to get a physical copy of the key as well. In other words, **anyone who puts a copy of their keys on social media with sufficient resolution has a security issue that they need to eliminate before it is too late**.

Figure 47. Cool key but...why am I seeing it?

Email Addresses / Domain Names

Knowing who owns a domain or a contact email can serve not only to identify you, but to look for data leaks related to that email (or the company to which the domain belongs). And of course, *how many people use the same key for several things? and/or does not use MFA?* Well, that is it... in **Seminar 1** and in **Lab 2** we saw tools to be able to see **who owns a domain** (`WHOIS` services) and look up this information. We also have the help of the search engines mentioned in the corresponding section of this document.

Received emails are also very interesting because they can be used as a content source for very interesting research. The following ***OSINT Dojo*** scheme (https://github.com/sinwindie/OSINT), is a very good summary of what can be obtained from an email you receive from someone (now you understand why having fake accounts that receive emails from the investigated targets is more important than you think 😊). It also does not hurt to look for pieces of the content on *Google* in case it was **generic emails** copied from a template (which is typical of programs like SET, as we will see later), or **known *scams***.

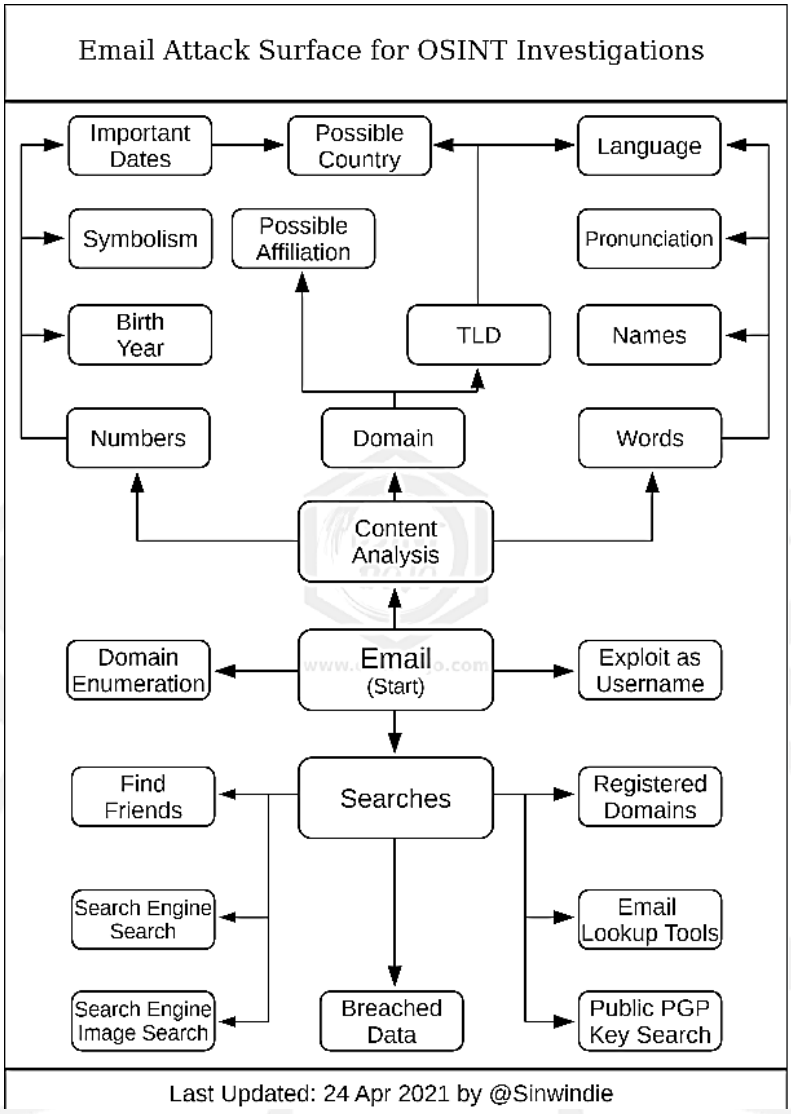


Figure 48. Email attack surface

Another alternative scheme you can follow is this one, made by Michael Bazzell (https://www.linkedin.com/in/michael-bazzell-a83572122/). Notice that in this diagram there are mentioned services that can help you in this task:

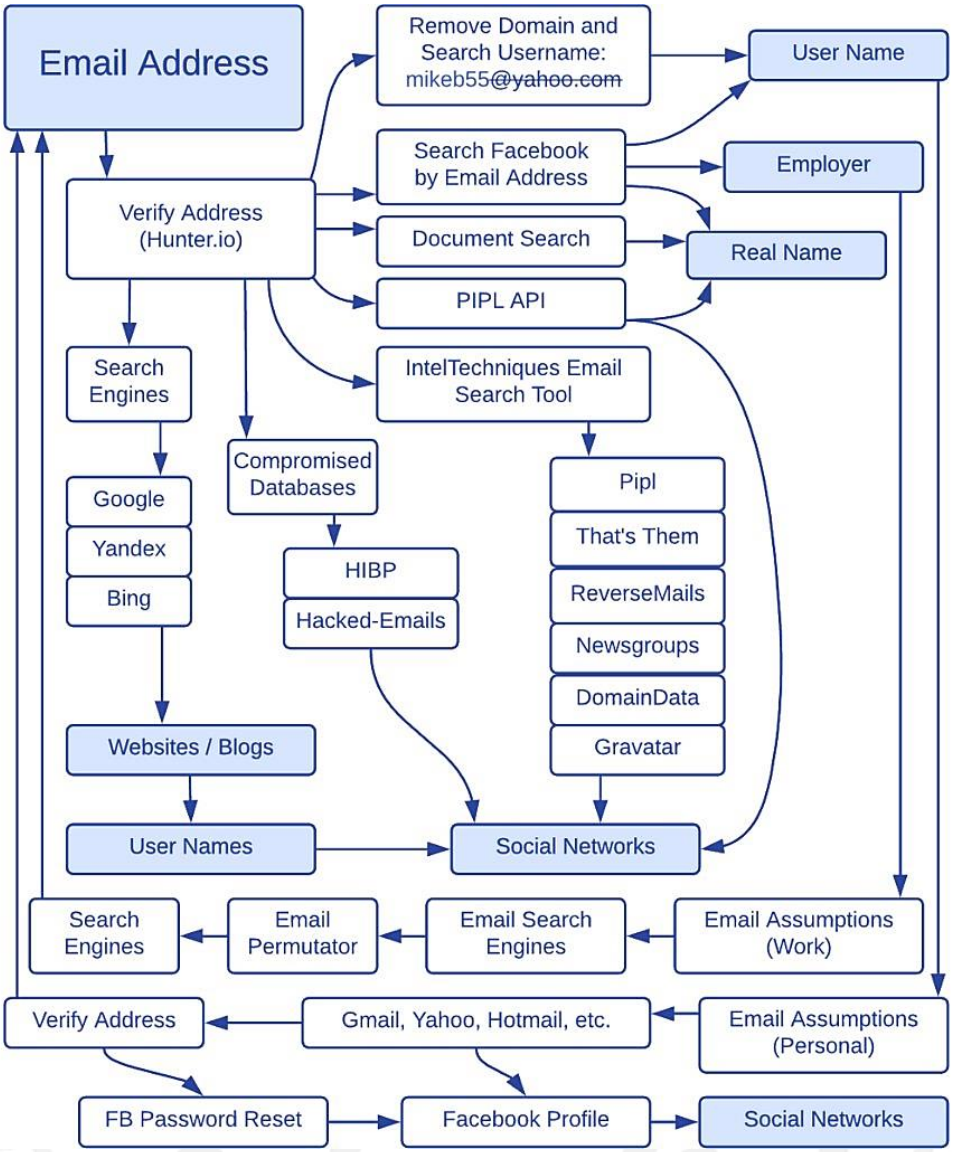


Figure 49. An email's attack surface, but following an alternate sequence

Tip to know how the owner of a Gmail account is

Sometimes we may need to know the **name of who is behind a *Gmail* account** that has written to us for some reason, and without alerting the issuer. It can be done easily by creating any document in *Google Documents* and sharing that document with that email account (but remembering to **uncheck the option to notify the user**). With this, *Google* will show us the name that we can then use to do more research. Please remember to unshare the document later, so you do not to leave any trace of this operation. This technique can be used, for example, to unmask who is behind *Gmail* accounts (although it may work with other equivalent providers, such as *Outlook*) that send us inappropriate emails.

NOTE: There is a similar technique that is done by sharing a *Google Calendar*, but this does notify the user that we have shared something with, whether we want it or not. ..

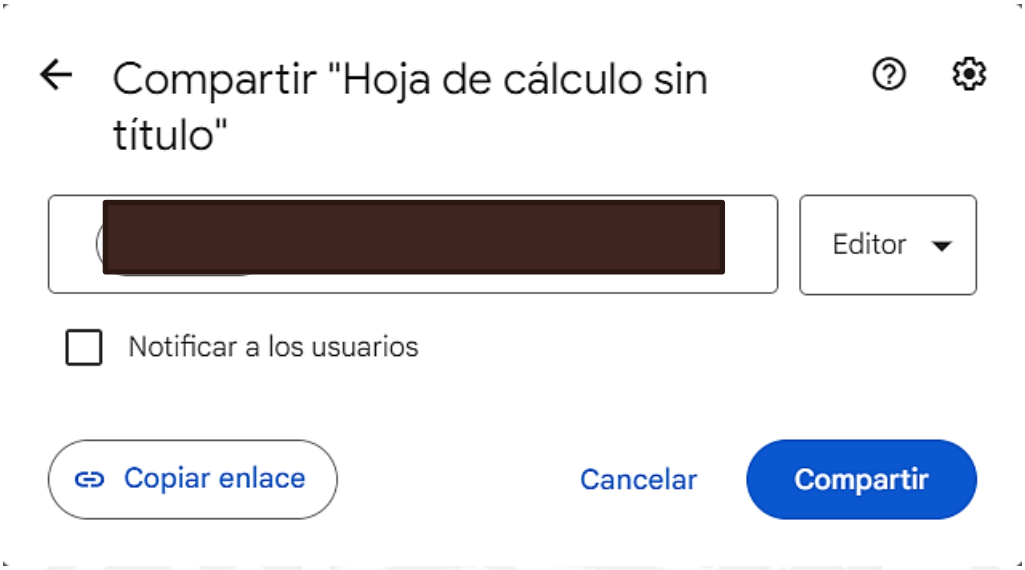


Figure 50. Sharing a document with an unknown Gmail account

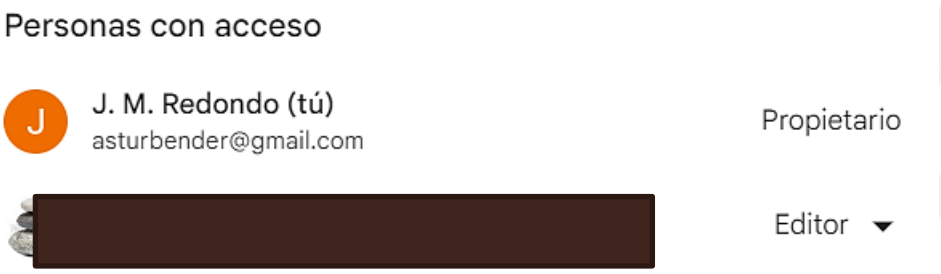


Figure 51. Name of the account owner in the list of users that have access to it

A similar trick, although it only works if the receiver has the settings configured to show the name, is to make a transfer for a very small amount (the minimum that PayPal *allows) to* a **PayPal account that may be associated with that name**. In some cases, you can reveal the name of the person behind an email address if you have it registered with *PayPal* when you create your account. More information: https://www.quora.com/Does-PayPal-show-your-full-name-to-the-sender

Social Media

As we said, **every photo you put of someone makes it a possible target to study**. But that is not all. Also think that if before you appeared in photos a lot with a person and now you no longer do it, you may be revealing a **distance / breakup** with that person, and therefore he/she may be upset with you. The same if you suddenly delete photos of yourself with that person, or from the person itself, unfollow them, or comment on improper things about them. There are ways that this person can be tricked into revealing valuable information about you, and it is easier if they are angry or upset with you (obviously, the same can happen with any of your friends...). It is a bit the same thing if you locate a **former employee** of a company (for example, using ***LinkedIn***): he/she may be more likely to reveal secrets of the former company, which are later needed for other types of attacks.

Figure 52. A person and their friends/study targets. Source: https://ciso2ciso.com/introduction-to-doxing-osint-methods-for-information-gathering-by-hadess/

Finally, never discard the possibility that the same person has a social network more closed than another, and that while in one you have hardly any relevant information, in another he/she is much more "talkative". To locate these alternative social networks, many times you simply have to look at the header of the person's network, since it gives that information directly, or use tools such as ***Project Sherlock*** (**Seminar 1**), which locates when **a person uses the same identifier in different social networks**, something that is very common. I myself have "uncovered" the hypocrisy of a couple of people who denied having done anything in one network while boasting of having done it in another, thinking that it was not possible to establish a relationship between them (something very frequent in people with a non-technical profile).

How do I manually study a social network?

The biggest problem of a social network is not that studying it is in itself difficult, it is getting an idea of the information it contains and how to extract it correctly, since there is a **large amount of information** and **doing it in an organized way can be really very complex**. Luckily, the ***OSINT Dojo*** repository (https://github.com/sinwindie/OSINT) has a series of schemes called "*attack surface*" that are excellent for doing this kind of research. Although the **repository contains many** more, we reproduce here the schemes of the networks that we have considered most important in order to do OSINT research so that you can use them more easily.

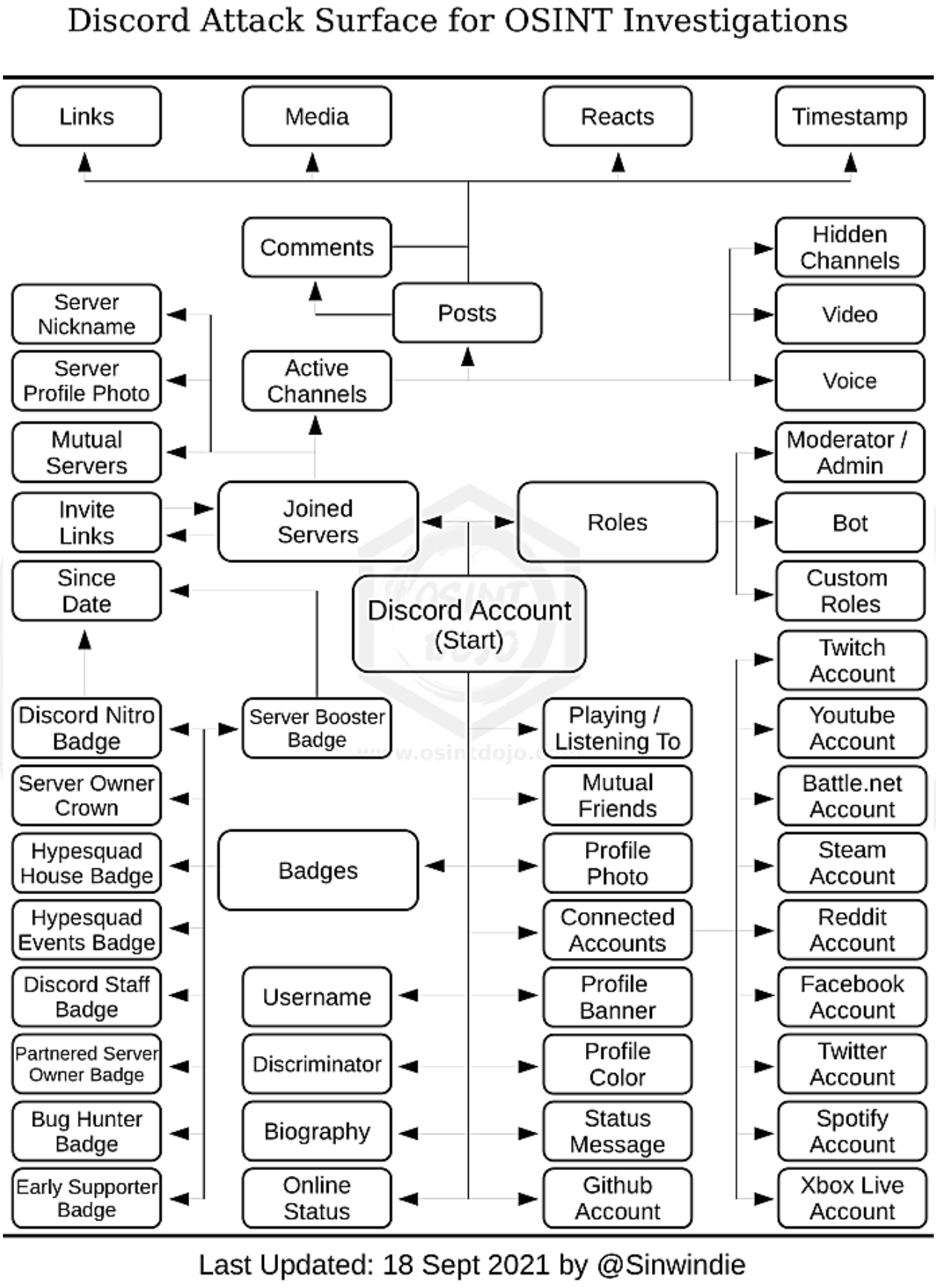


Figure 53. *Discord* attack surface

YouTube

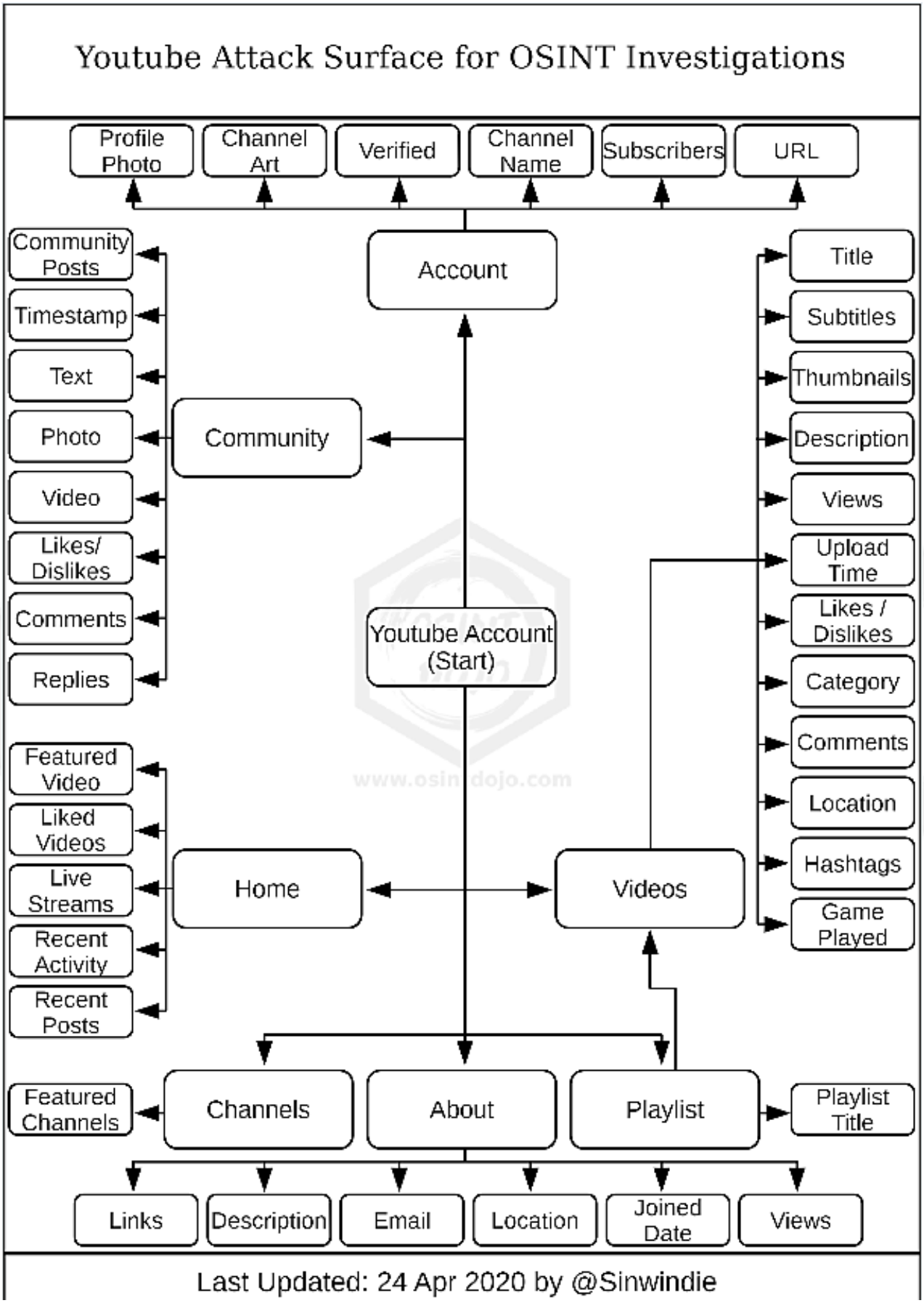


Figure 54. *YouTube* attack surface

Twitter (Does somebody call it X? 😏)

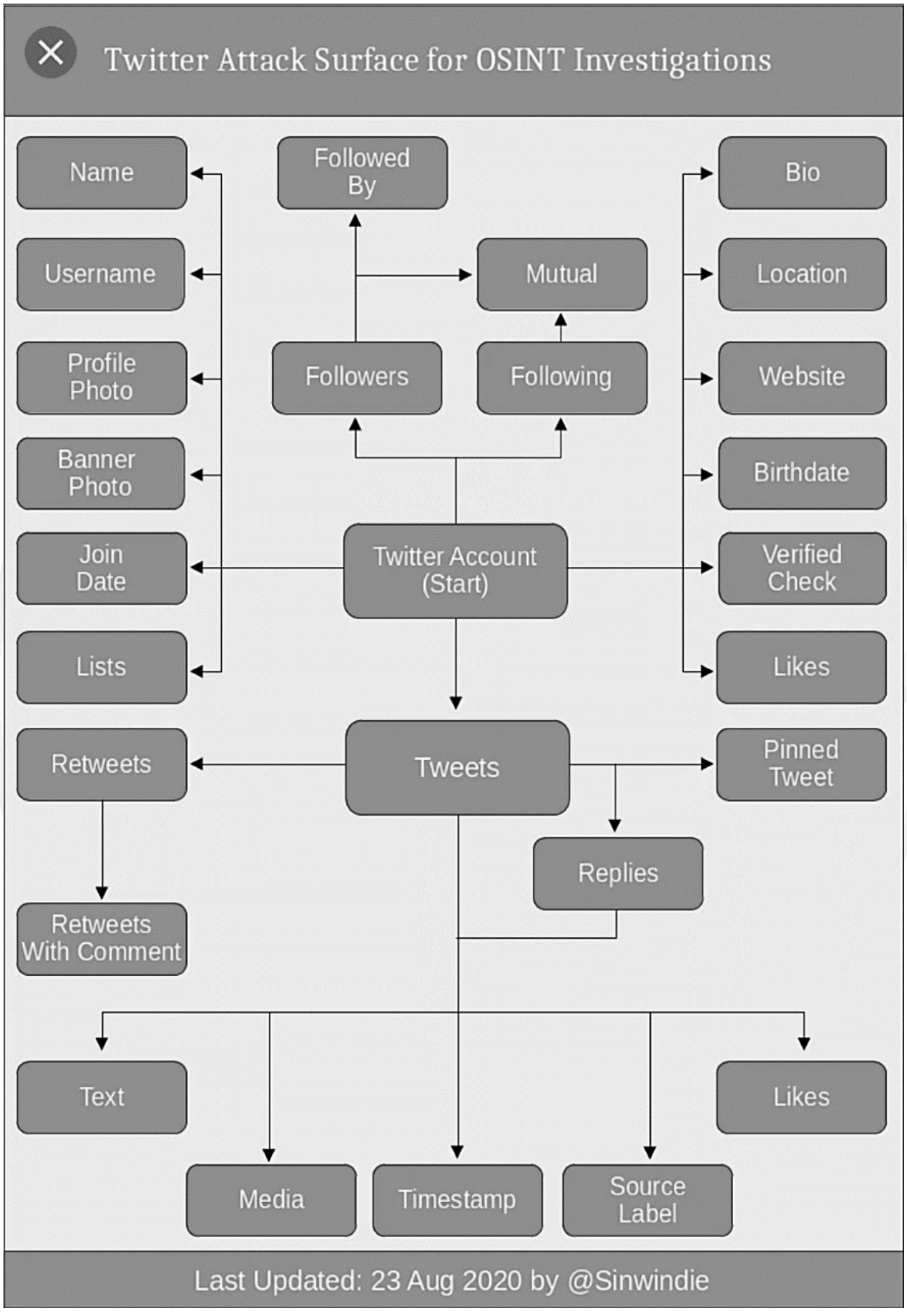


Figure 55. *Twitter* attack surface

Instagram

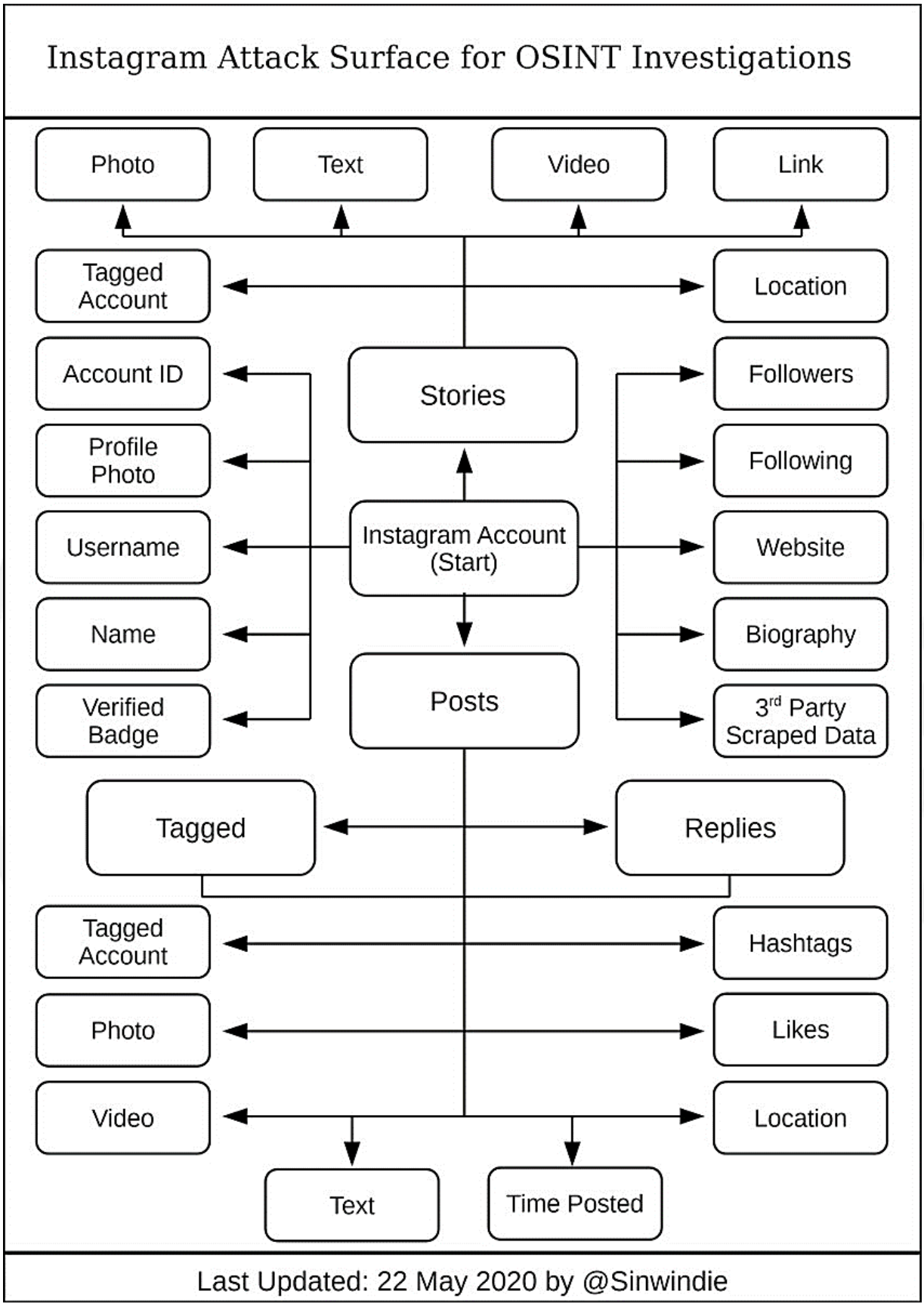


Figure 56. *Instagram* attack surface

LinkedIn

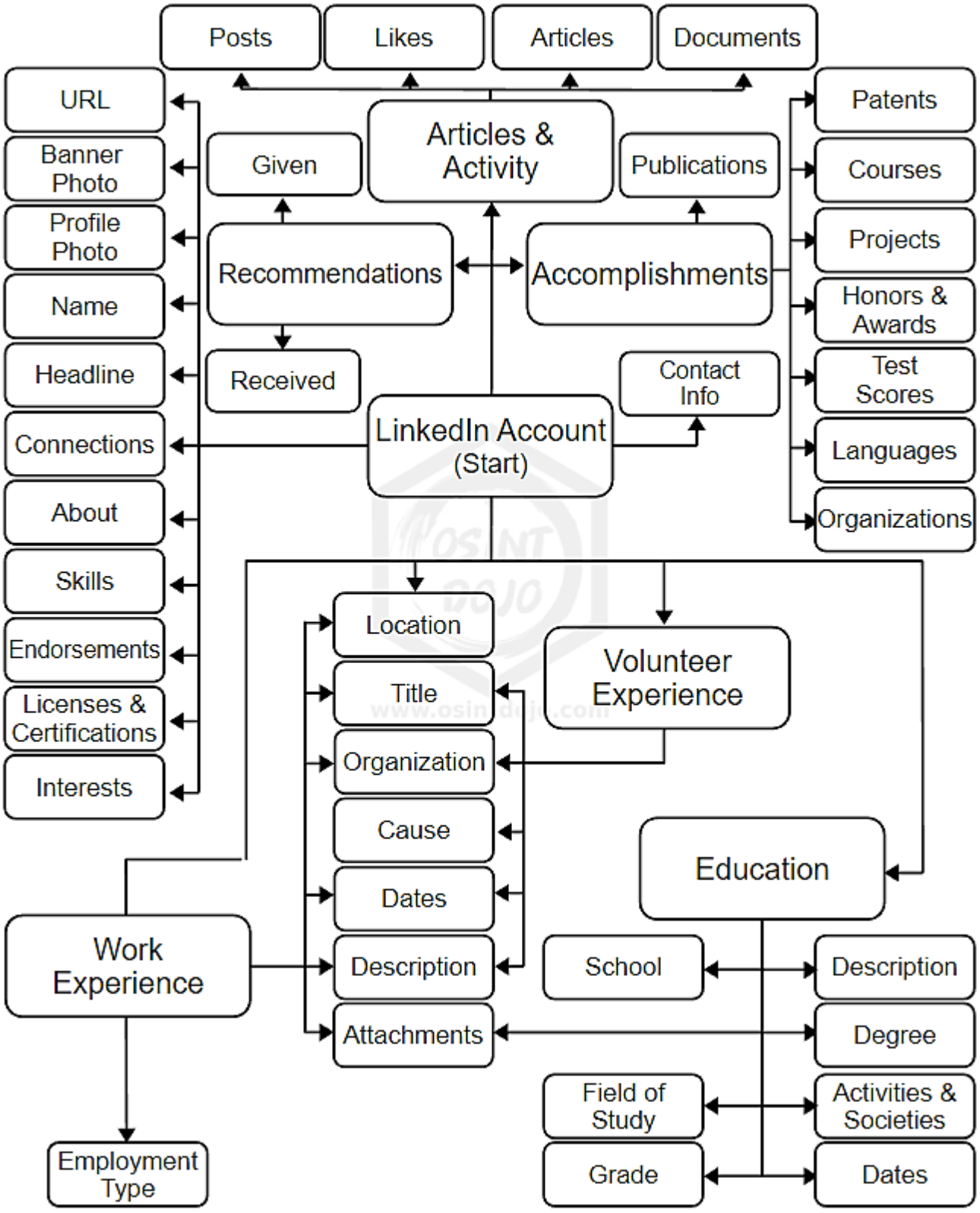


Figure 57. *LinkedIn* attack surface

Snapchat

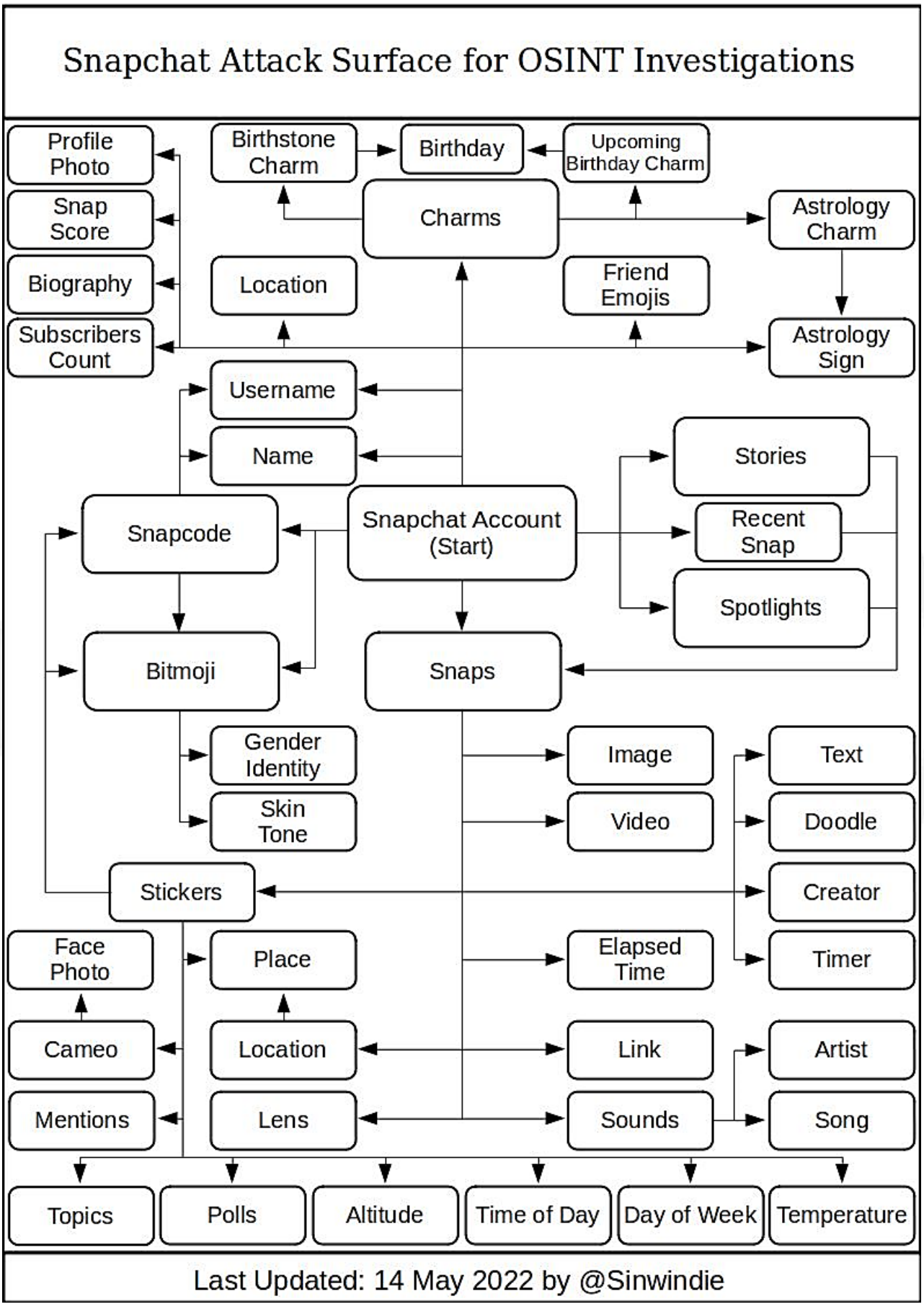


Figure 58. *Snapchat* attack surface

TikTok

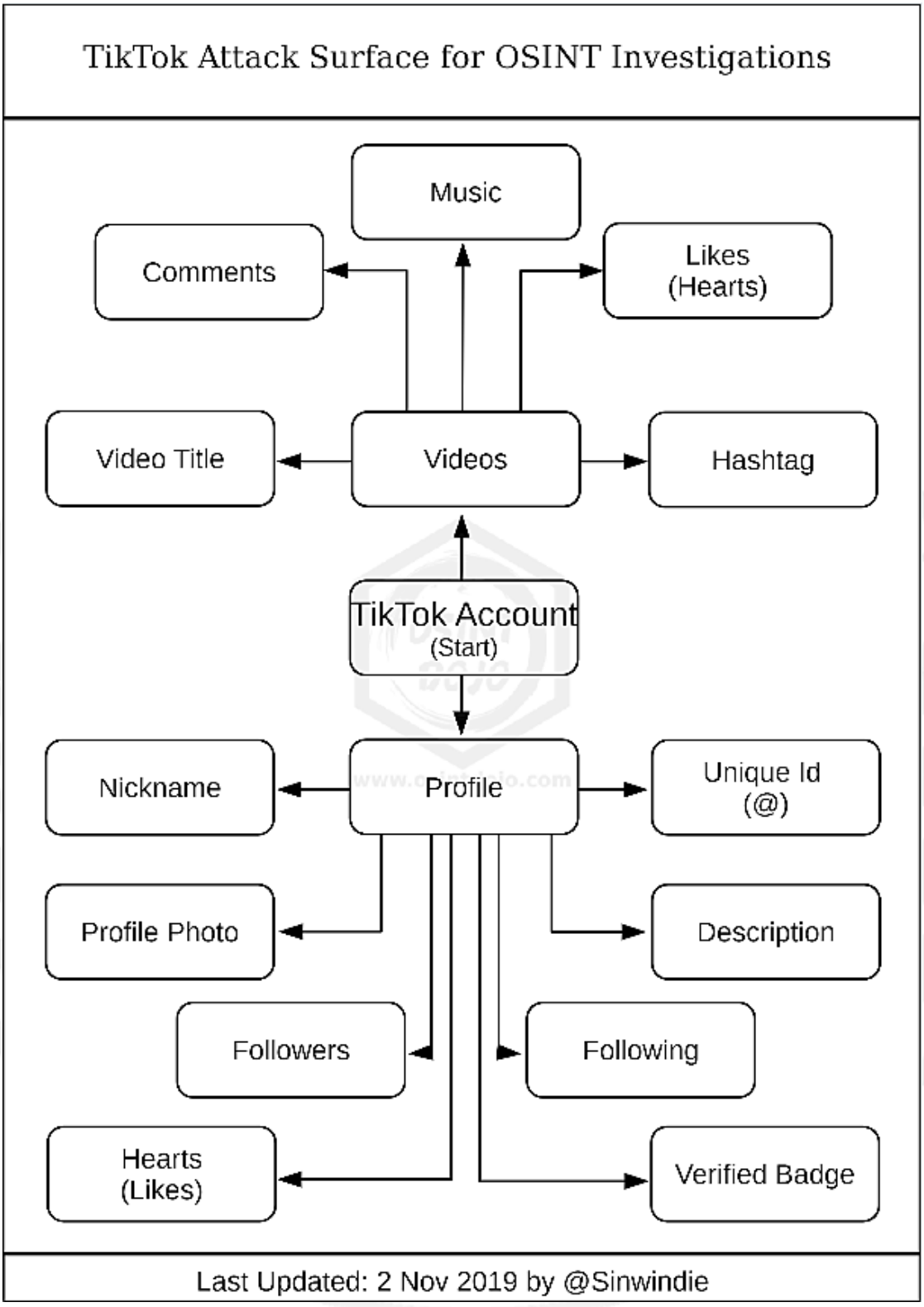


Figure 59. *TikTok* attack surface

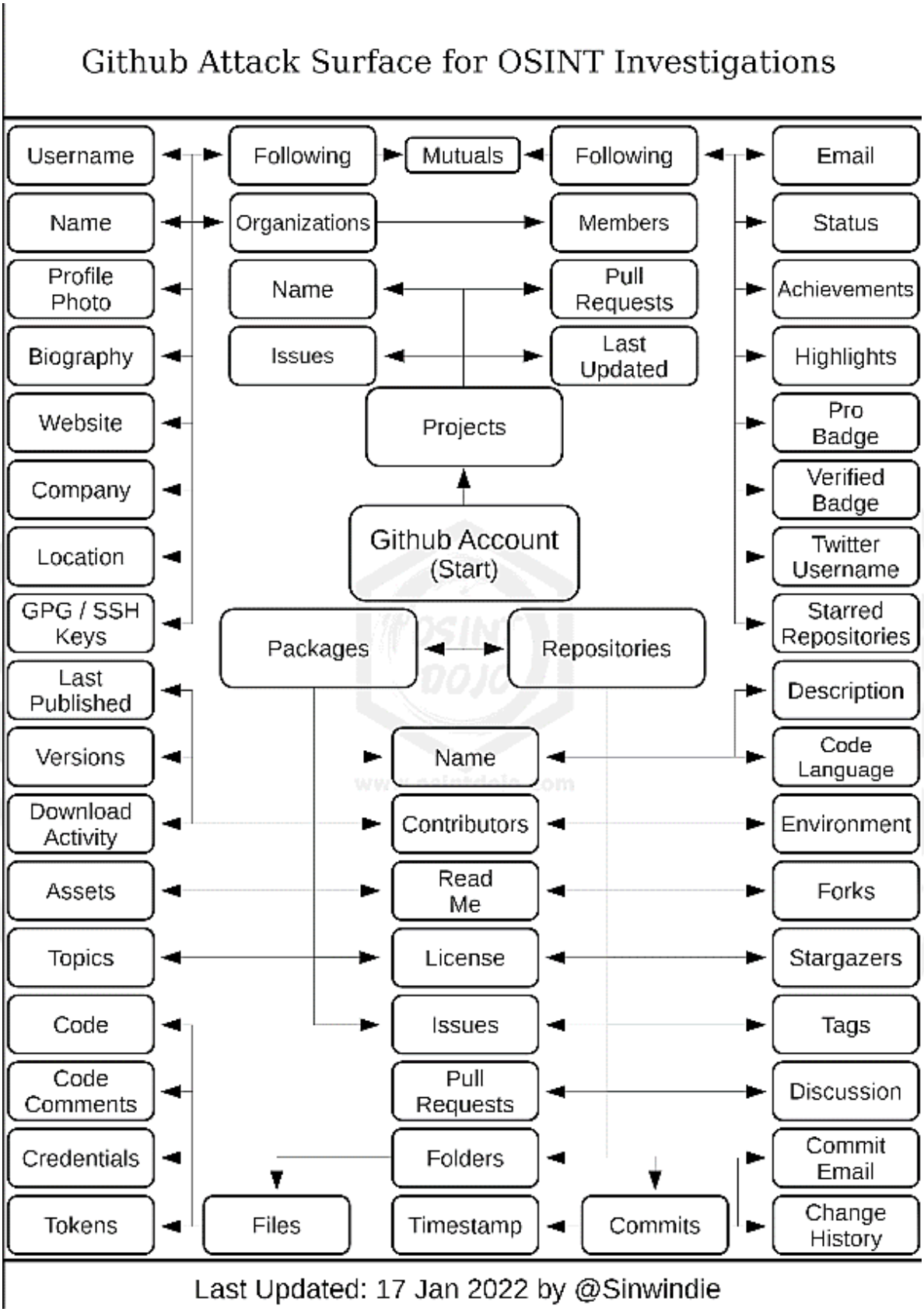


Figure 60. *GitHub* attack surface

What if I want to study them with tools?

In that case, you are going to have an advantage and a disadvantage. The advantage is that there are usually many tools that work on different social networks at any given time, and many of them are specialized in

obtaining specific information from each of them (images, contacts, relationships...). The problem with these tools can be summed up in a number of points:

- **Anonymous people maintains them,** and you should trust that their intentions **are not malicious** (i.e., that using a tool does not install *malware*). You have to use an isolated execution environment, look at the reputation of the author and the tool, etc.
- **Social networks usually do not like these kinds of tools**, so they could stop working without warning. The most representative case is *Twitter* and the closure of its API to convert it into a paid one.
- In line with the above**, authors can stop maintaining them at any** time (after all, they do not get paid for it), and a tool that once gave you functionality can be useless in no time...

If you are aware of this, you can find many tools as soon as you search. Not only in the ***Ciberpatrulla*** repository mentioned above, but also *in* ***Social Media OSINT***, specialized in tools of this kind: https://github.com/cqcore/Social-Media-OSINT

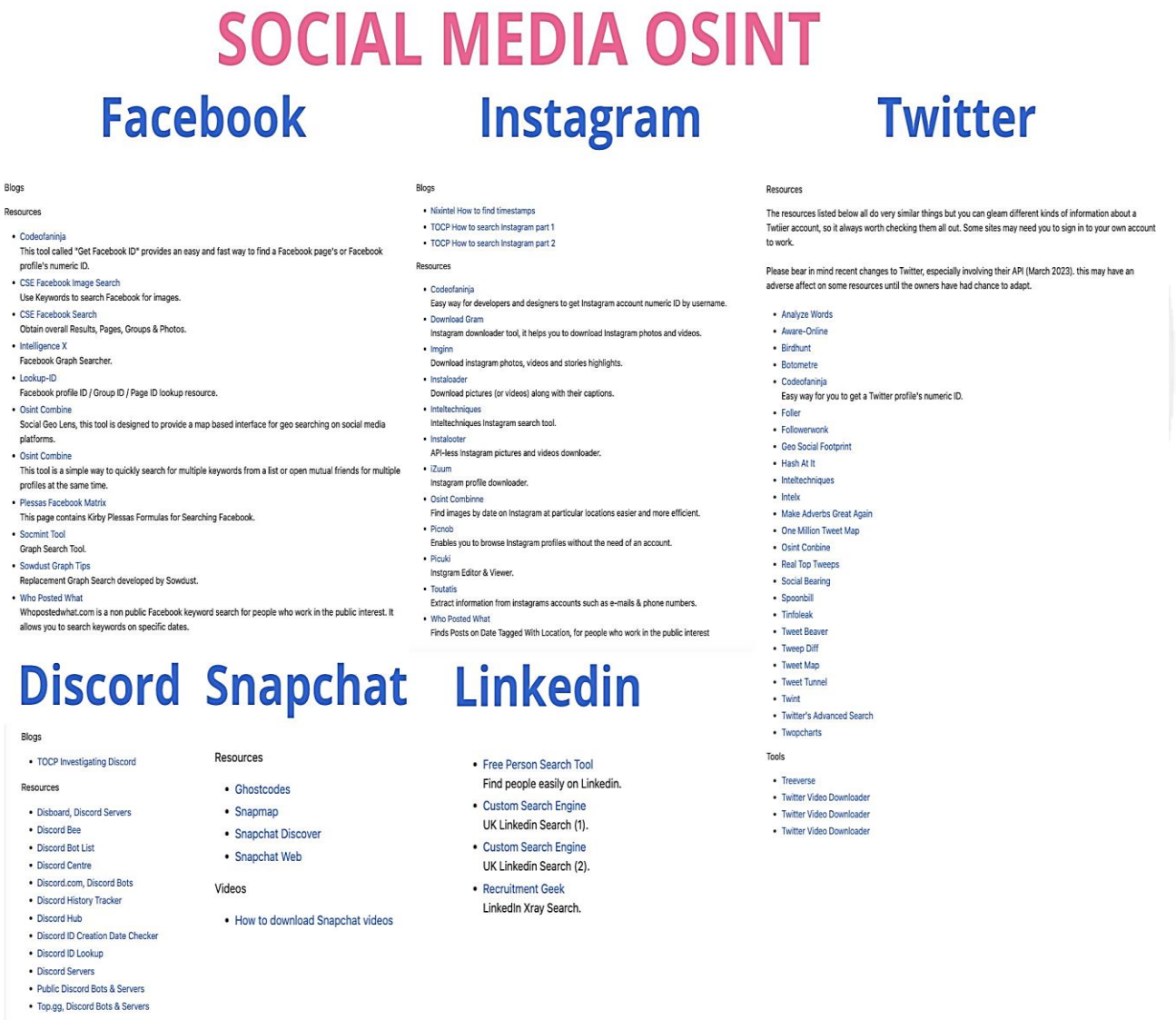


Figure 61. Summary of the contents of the Social Media OSINT repository. Source: https://twitter.com/cyb_detective/status/1647588776762388480

About "B Account" on social media

There are people who use alternative accounts to their main one on social networks. Many do it to maintain anonymity, so that in "Account B" they can **say or give their opinion things that, if done on their official account, could cause problems for certain groups,** lose reputation or any other effect. Or you simply want to exercise your right to be anonymous on the Internet. That is a separate issue from the one we deal with in this document, which is to know when an account is a "B Account" of another because in this way we will know the identity of the person behind it.

While this is not an exact science, there are several clues that give away a "B Account" of this kind that you can use to see if you are looking at one or not:

- Account A has a lot of followers but follows few people. Among them, there is the candidate "Account B" that you want to investigate.
- "Account B" always reposts "Account A" no matter what it puts in, and it really has few interactions of its own.
- "Account B" always defends all the positions of A, even very vehemently.
- "Account B" is dedicated to insulting and disrespecting anyone who has an opinion against A or refutes any of its arguments.

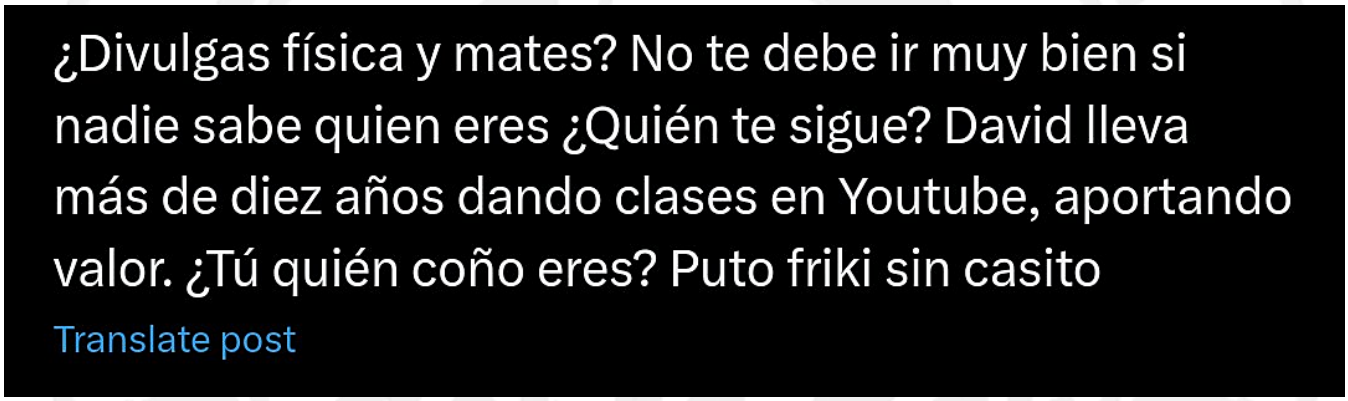


Figure 62. This post is supposedly from the B account of a well-known communicator

People search networks

We have talked about social networks, which are also a search network for people in general. Also, with "typical" search engines you can achieve the same effect. If that is not enough, there are **search engines of people** who are expressly dedicated to that, and that we can use as a complement.

Some of them are mentioned in this article, although they only allow people from the US to search, and probably your goal lives in another country: https://www.lifewire.com/search-engines-that-top-the-web-3482269. An example of a search engine precisely **for people in the US** (you may already notice that privacy in the US is not taken care of as much as in Europe) is ***Intelius***: https://www.intelius.com/.

If a company needs to do something like this in a serious way, it should consider professional payment options, such as ***Pipl***: https://pipl.com/product/pipl-search. As this is not our case, in the course we can use the functionality that *Facebook* or (better) *LinkedIn* has for it, as shown in the referenced article.

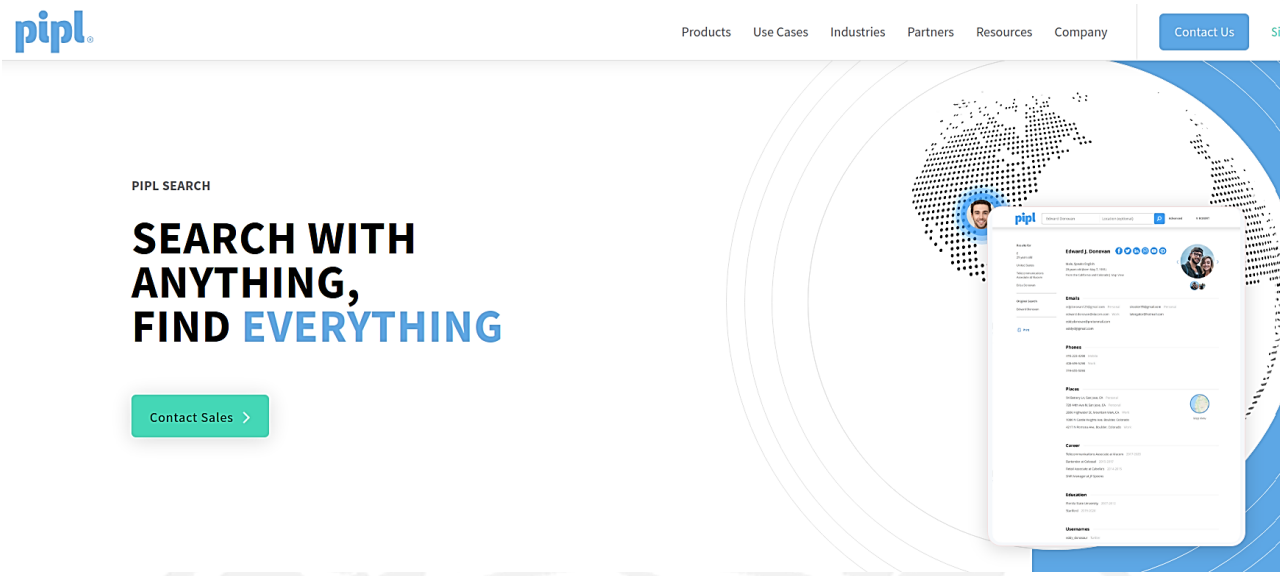


Figure 63. *Pipl* People Search Engine and API

If using web search engines is not your thing, there are also **tools that can be installed on your PC that do that kind of job**. Here are a couple of them:

- ***Mr. Holmes*** (https://github.com/Lucksi/Mr.Holmes): An application that searches for domain information, username, and phone numbers with various techniques.
- ***Alfred*** (https://github.com/Alfredredbird/alfred): A tool for searching for username information on the Internet.

Dating Apps

Although this may seem like a joke, it is not at all. Once you have basic information about a person, the probability that you can find them on a ***Tinder***-style dating app, ***Grindr***, etc. is not zero. If, in addition, the person who investigates creates a false identity to register in this network, as we saw in a previous section, it can be, in the hands of someone specially trained in how to talk to other people to try to extract information, **a very powerful way to find out data or secrets that make other attacks possible**. Obviously, this is an investigation that requires "engagement" with the objective, **so it is outside the scope of this course**, but again it is commented not for what you can do, but just in case some day you are the ones who are on the side of the investigated, and not the investigator. Doing this type of investigation legally is the job of law enforcement.

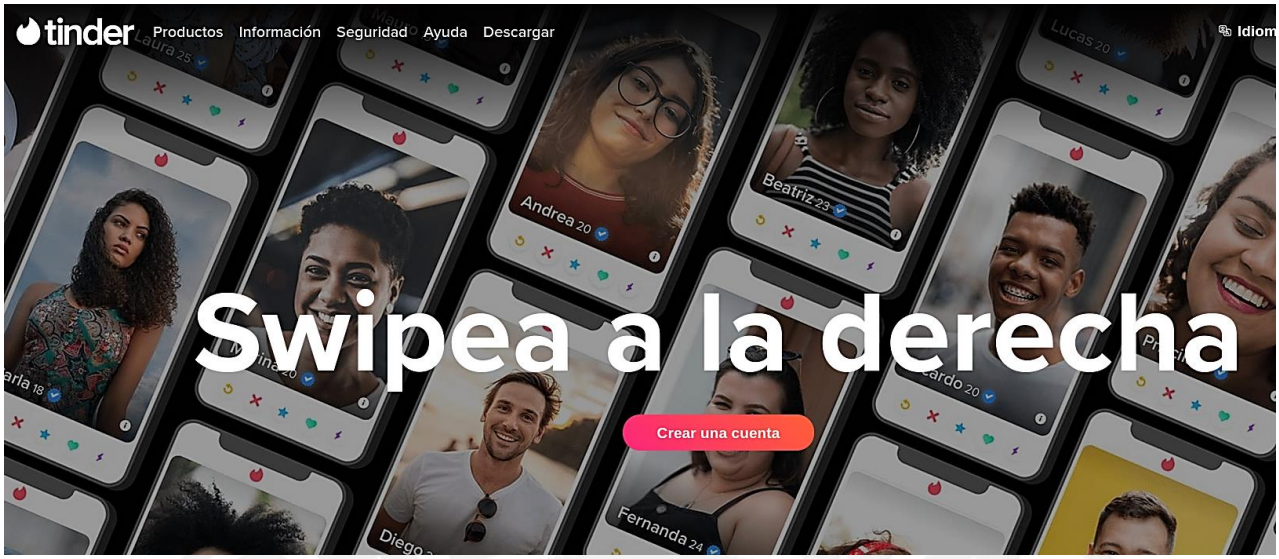


Figure 64. With a photo, a name and an approximate geographical location of a target, sites like Tinder can be a point from which to get more information from those investigated.

Phone numbers

In 2012, one of the largest legal leaks of people's data died: **the white pages**. A huge book that came to many houses for free, and that listed all the telephones in Spain, sorted by surnames and showing the address of each person. Obviously, if it continues to exist, it will facilitate the work of investigation and obtaining data to make spear phishing attacks or similar. Which does not mean that maybe in your house there is one of them forgotten, and it turns out that you locate the phone (and the address) of someone who has not changed it in these years ...

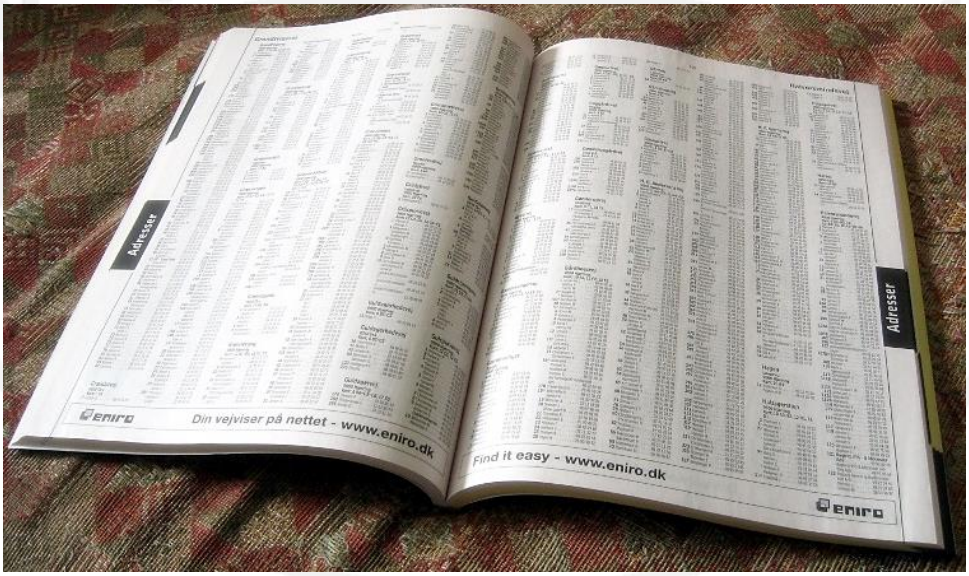


Figure 65. A physical White Pages guide, which is like printing a modern data leak. The 2000s were wild in many ways... Source: Wikipedia

This is the past, as well as the **yellow pages** (a national phone book of **companies and businesses** only), which also stopped printing 9 years after the white ones. But what happens in this case is that there is an online version available here: https://www.paginasamarillas.es/, which makes it possible to find more business information, if they are our study objectives.

Figure 66. Yellow pages searching for business information

However, if we are looking for the number of someone from Spain or any EU country we can only rely on finding the phone filtered on the Internet by accident (or not, offered by the person directly), since **telephone searches**, whether direct (from person to phone) or reverse (from phone to person) **are prohibited by the data protection law**. It is only possible to use websites that collect phone numbers that are identified and evaluated by individuals, which is usually done when they correspond to businesses or phones used for aggressive or inconvenient telemarketing practices. One of the most well-known pages for that is this: https://www.listaspam.com/

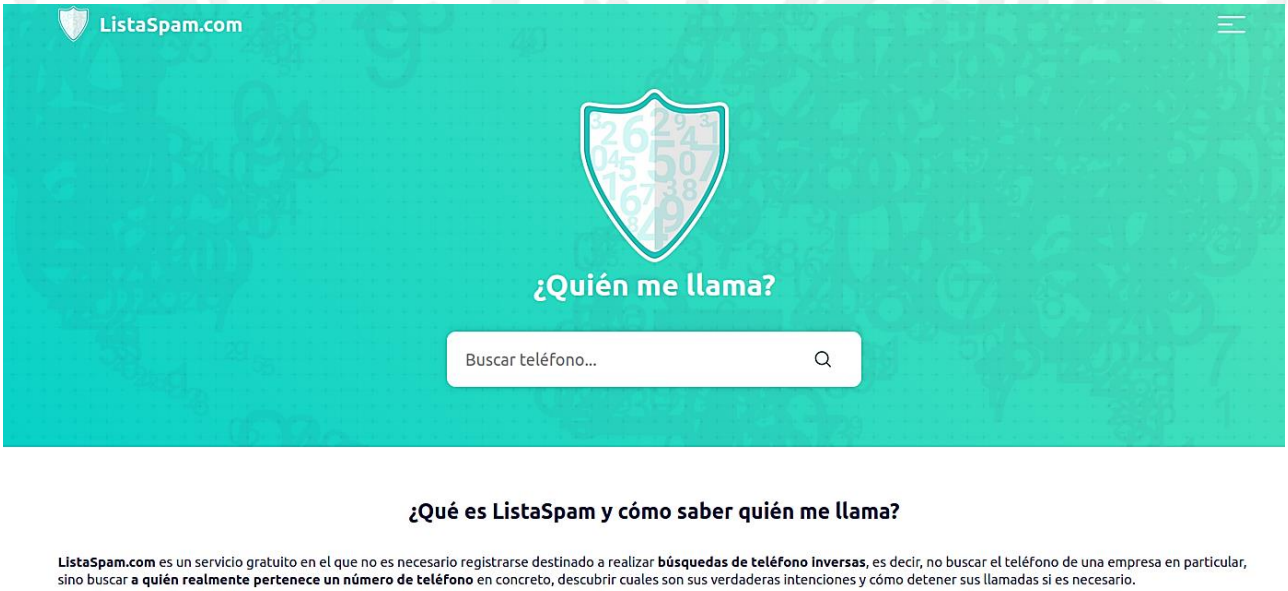


Figure 67. *ListaSpam* and the like at least lets you know if a target is doing aggressive business practices or considered annoying

It is also true that some companies or organizations (private or public) could keep a **directory of their employees** that can give more information about them, and that is worth consulting. This is for example the

case of the University of Oviedo, which allows anonymous (limited) searches: https://sies.uniovi.es/paginasblancas/web/busqueda-anonima/personal.faces

Figure 68. White pages of the University of Oviedo employees

It should be said that these restrictions **do not apply in other countries outside the EU**, such as the US, so if the person you are looking for is from there, you can do a reverse search with services such as ***Spokeo***: https://www.spokeo.com/

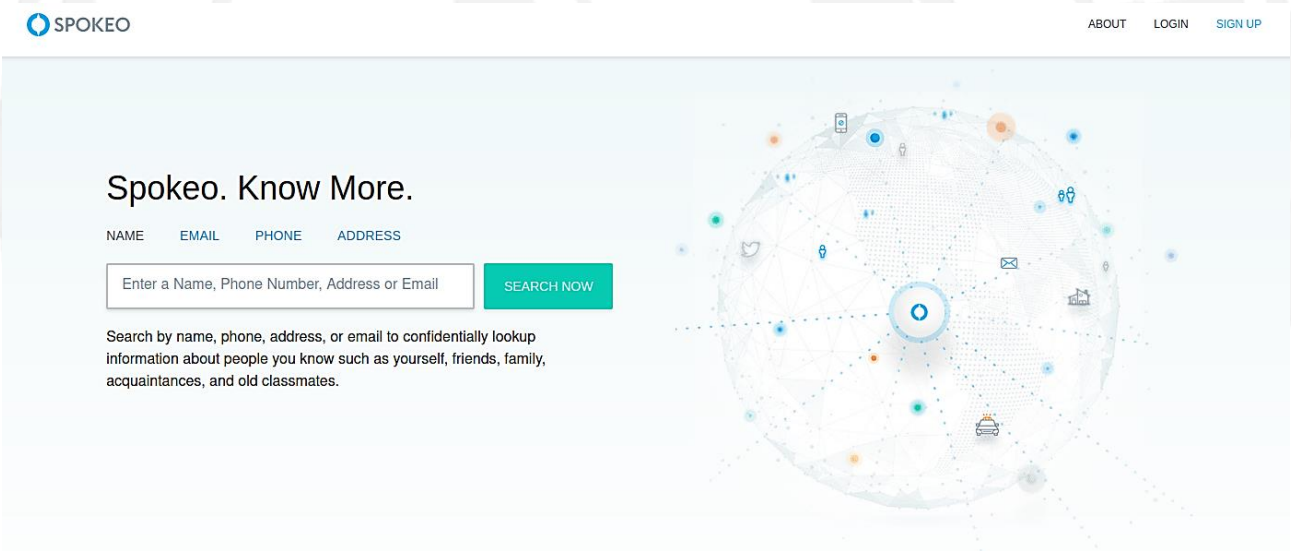


Figure 69. In the US, the issue of the privacy of phone numbers is on another level. There are even criminal record seekers as well....

Partial phone number from PayPal accounts

Sometimes some services show us part of an account owner's phone number when we do certain operations, **such as remembering passwords or similar**. Right now, it is what happens with *PayPal*: it is possible to know **5 digits of the phone number** of the owner of any account (and its prefix) or its email address. Sometimes even part of their credit card and card type. This information can be used to facilitate searches that locate data of the target or, in combination with other digits that reveal other services before a similar operation, even have all or a great part of something as private as your phone or card. **It is advisable however to do a test with our own account** of the service in case the owner receives a notice that these tests are being done (in the case of *PayPal*, in March 2023 he did not do it). Or imagine that the objective is

someone you know and have in your telephone contacts. Checking these numbers against them will reduce the amount of possible targets

Figure 70. *PayPal* revealing too much information of an account holder

You may think this incomplete information is not worth much, and the truth is that using it in isolation is not worth it. But now imagine that person uses *PayPal* for something that harms you and you suspect that it is an acquaintance: with this data it is very easy that there is only one person in your contact list who fits with them!

Other ways to get the owner of a phone number

If you have a phone number but have no idea who may be behind it, there are two more things you can do to see if they can give you more information about it:

- **Send a new message to the phone via *Telegram*** adding a new contact (e.g. https://www.makeuseof.com/how-to-add-a-new-contact-on-telegram/). If that person has a moderately public profile, we will be able to know things about him or her, such as a photo, real name, and username. With that username, if necessary, we can search through different social networks if there is someone who uses the same name and, although it is not 100% safe, it is very likely that with that we will get more information about the person we are looking for.
- **Start the process of sending a *Bizum*** for a ridiculous amount to the phone number (e.g. 0.50 euros) at our bank. We will be able to know at least part of the recipient's name (depending on how it is configured) **without having to confirm the money transfer**. Generally, before you make the transfer, you should see a confirmation with the name of the recipient.

In any case, you can follow this general scheme when it comes to finding phone number information:

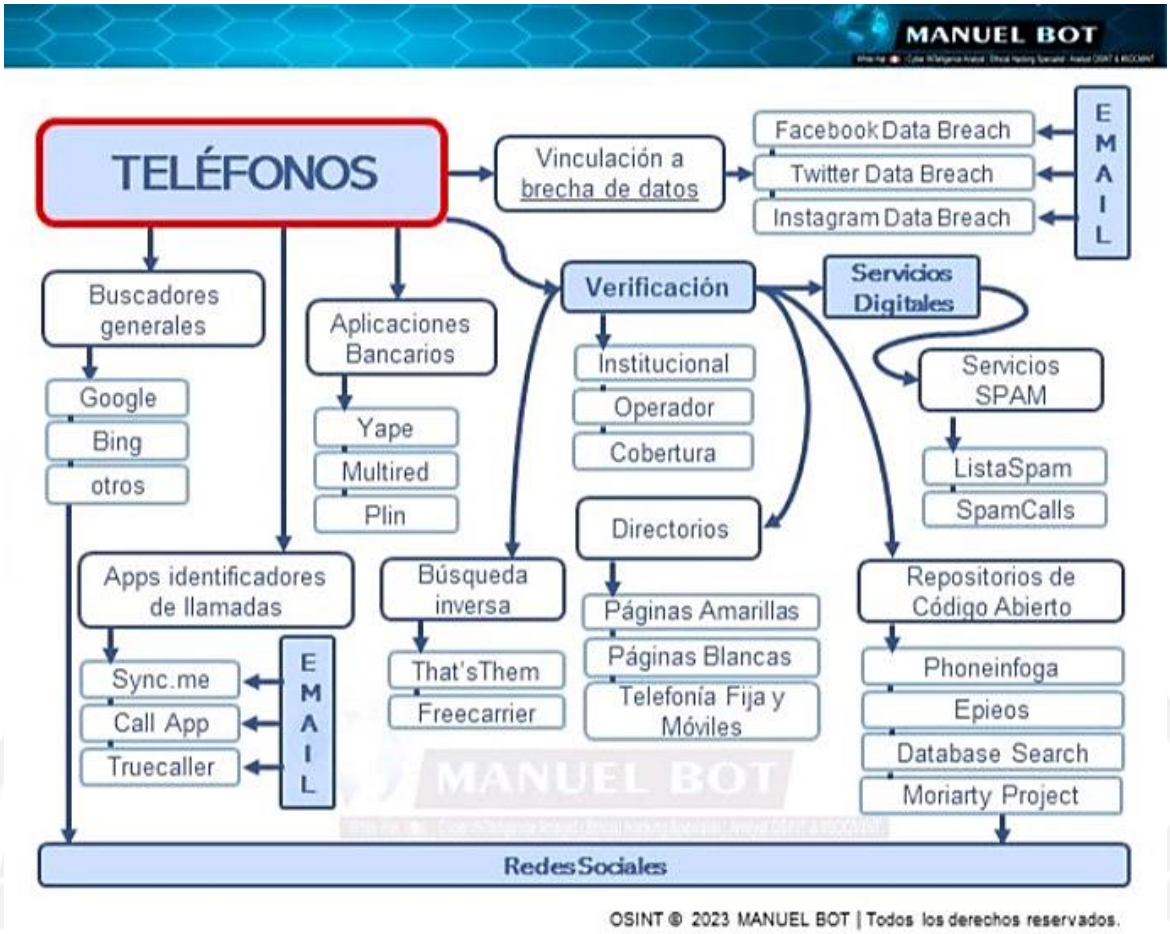


Figure 71. Where to find information about a phone number, according to Manuel Travezaño (https://www.linkedin.com/posts/manuelbot59_osint-workflow-saberespoder-activity-7142187376654643200-ws0-/?originalSubdomain=es)

Public Records

There are several public places where it is possible to find several information about people, although it should be noted that it is possible that access to it **has a cost**, **a plausible reason to do so has to be provided** and / or **the owner is alerted that this search has been made**. Before using these services, you have to look very carefully at the conditions in which the usage occurs, since it can very easily fall into the category of "engagement" that **we are NOT looking to do in this course**. Obviously, if in the future any of these public records were leaked by an attack, the confidential information of an entire country would be seriously compromised (there is not much talk about the leak of the SEPE at the beginning of the pandemic, but it is scary to think what could have been exfiltrated then ...).

- **Mercantile Registry** (https://www.registradores.org/el-colegio/registro-mercantil): Procedures to create a company in Spain, Public Limited Companies, companies with contracts and state purchases, directory of business records, taxes, economic and financial data, legal and economic data of companies and individuals, who is the representative of a company or society ... It can be searched by name or registration number. Basic information can also be obtained free of charge from the databases of private companies such as ***Axesor***: https://www.axesor.es/

Figure 72. Axesor website with free basic information of any company in Spain

- **Land Registry** (https://www.registradores.org/el-colegio/registro-de-la-propiedad): Public data of properties and real estate to know who the owners are.
- **Civil Registry** (https://www.mjusticia.gob.es/es/ciudadania/estado-civil/registro-civil) for births, marriages, divorces, deaths, date of birth, etc.
- **Electoral Census**: Checking whether or not someone is in the census implies prior authentication (so they only let you consult those data about yourself), so it is not viable to do it anonymously for a third party from the website enabled to do it by the corresponding ministry: https://sede.ine.gob.es/ss/Satellite?c=SETramite_C&cid=1254734728608&lang=es_ES&p=1254734719739&pagename=SedeElectronica%2FSELayout. However, this type of lists is usually published on paper in the town halls at election times, and these physical copies can be used to extract information from other people without anyone noticing it.
- **Vehicles** (Queries by license plate number): As said above, it is possible, upon payment, authentication, and providing a valid reason, to know who owns a car, if they have up to date the payment of taxes, fines and Infractions by license plate number or name in the Ministry of Transport. It may seem excessive, but it is something necessary if you are going to buy a second-hand car, to ensure that it is "free of charges", so the data is accessible.
- **Cadastre** (https://www.sedecatastro.gob.es): It is a way to obtain information on farms, maps, plans, real estate, real estate, credits, purchases, sales, loans, liens, mortgages of farms and condominiums. By requesting and paying a **simple note** in the cadastre you can find out who owns a property and where it is located, or even see the plans of a house. Free of charge and via the Cadastre web you can search for **satellite maps**.

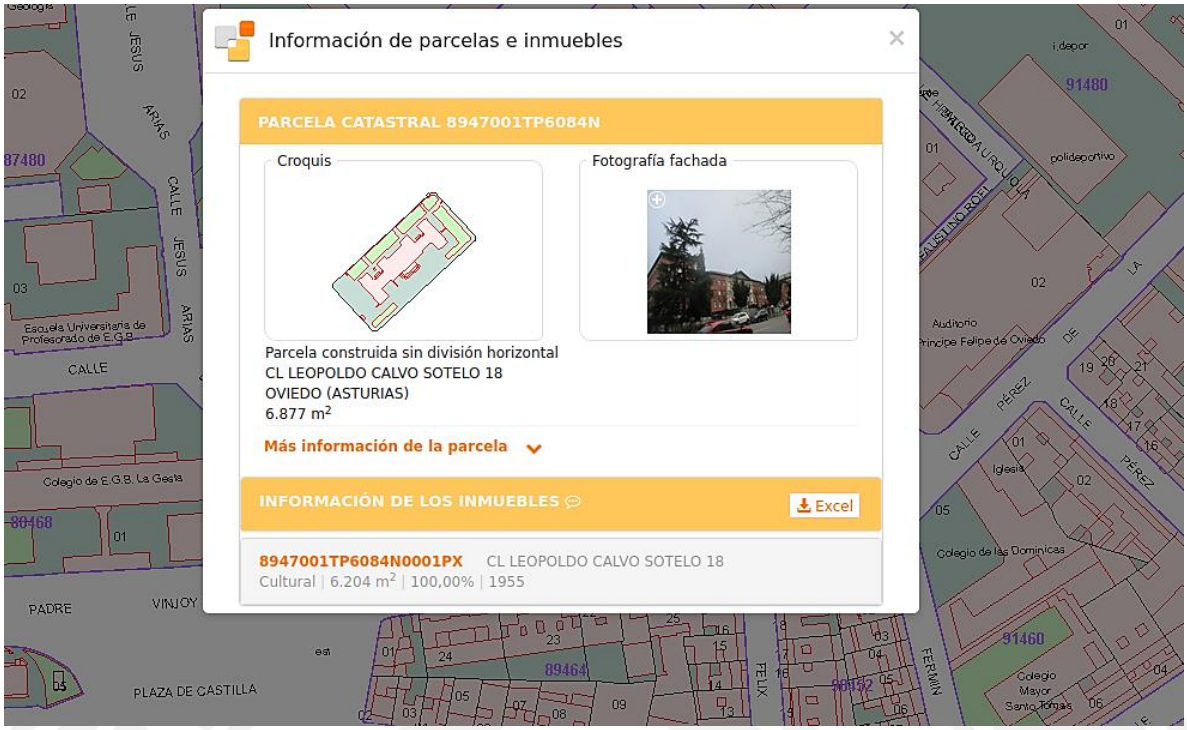


Figure 73. Satellite map of the Spanish cadastre

To access the data of a property the first thing we need is the **cadastral reference**. But in the map of the previous image, we can easily access that code (20 characters), so it should not be a problem. We should not confuse this number with the **IDUFIR** (IDentificador Único de FInca Registral) of a property, since these terms are not related to each other. This document is also used to make the valuation of a home in free online portals. With this code we can request the simple note we talked about before, either digitally or in person at the cadastre offices. A request costs between 3€ and 10€ and is not immediate (it can take about 3 days). In this simple note will contain the data of the person who owns the property and the possible charges that it has, which again is necessary to buy a second-hand property with the certainty that it has no charges. Other search options include:

- **By registration data**. For this type of search, it is necessary to have the registration number of the farm, and the section and ordinal of the sub-farm.
- **Through IDUFIR**. As the name suggests, it allows you to find a specific farm within the national territory.
- **Per owner**. For this we need your full or partial name, or your NIF (see next point).
- **With other data**. This type of consultation is used for when we only have the location of the farm.
- **By geolocation**. From the coordinates of the selected point in the viewer, we can identify the IDUFIR associated with it, provided that the farm has been geolocated in the registry.

- **DNI / NIF**: The ID / NIF number of someone can be surprisingly easy to find since it can appear directly listed on the Internet, often without the person being aware of it. If we only have the DNI, the corresponding letter can be calculated by an algorithm that can be easily found on the Internet. To locate DNI / NIFs it is enough to do very simple searches on *Google* such as `"<full name and surname of the person>" DNI` or `"<surnames, full name of the person>" DNI`

  It is common, for example, for people **to post their resumes on the Internet / *LinkedIn*** (or similar networks), including a lot of personal and work information (even more than they give on social networks), and that we can use to complete the profile of the researched person. Among this information it is common to find the ID number, photos or even addresses. This is common in researchers or university professors, where creating a CVN (***Standardized Curriculum Vitae***) and

exposing it to the public is very common to be able to access projects or other academic uses. In these documents a large amount of data of the person and / or people and companies related to it is usually given (research projects carried out with companies, thesis supervisions...), which can be easily used to "pull the thread" of our research.

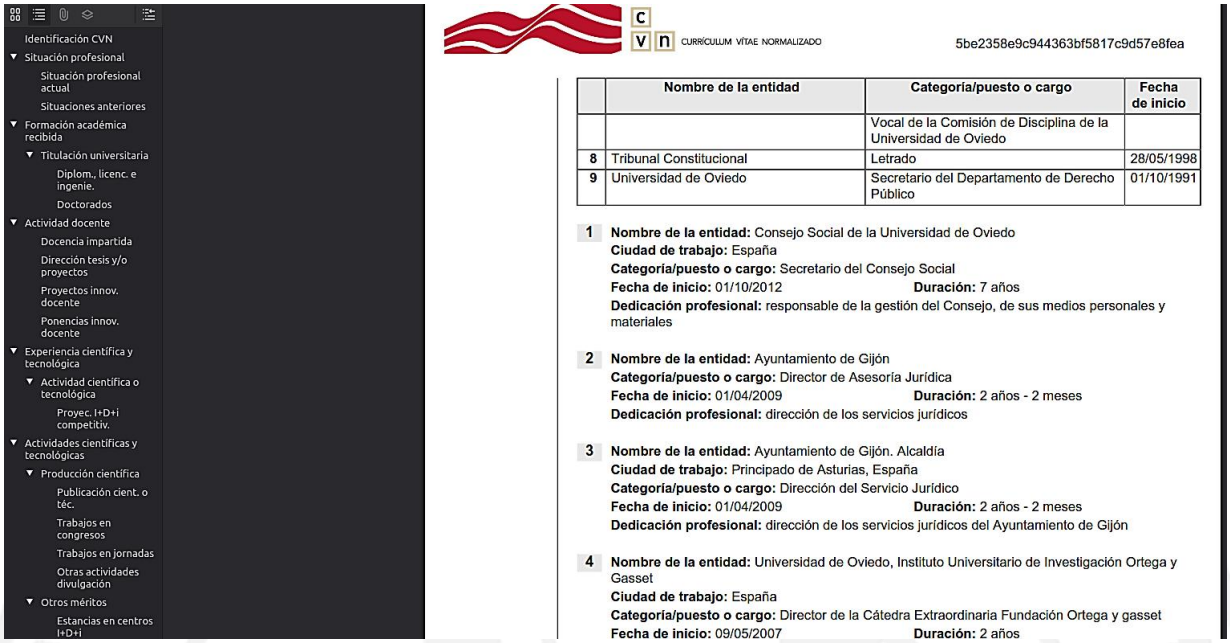
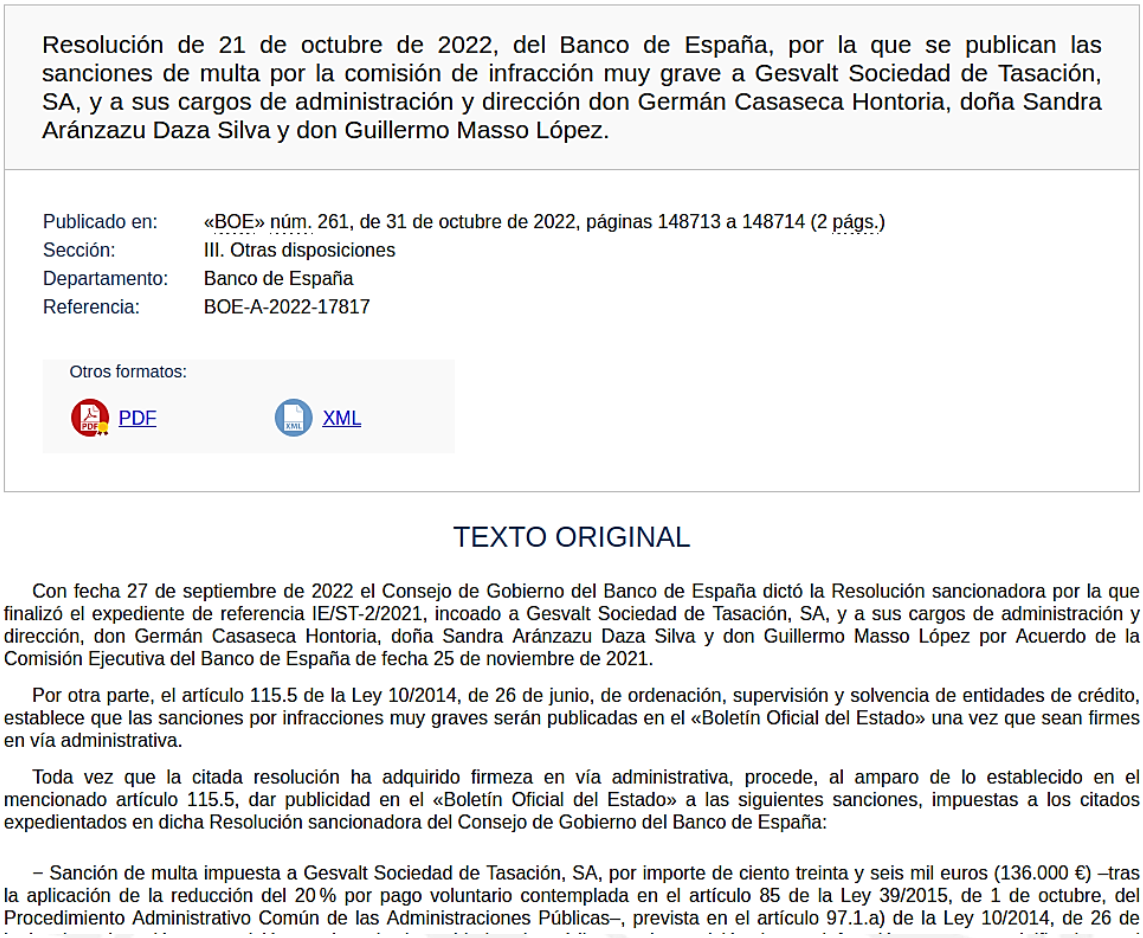


Figure 74. Public CVN of a member of the University of Oviedo

You should also not rule out the presence of **old "forgotten" documents** on the network where this information is exposed because they have not been removed from public access, either through direct searches such as the ***Wayback Machine***. Not so long ago, privacy on the Internet was not a concern and there were almost no measures against it. That information can still be there, and not even the owner himself could know. I have even managed to locate my university marks 15 years later...

- **BOE / or regional bulletins such as the BOPA**: If someone has been **fined**, or **has not paid a traffic or other fine**, it is possible that this will be published in the BOE or another regional bulletin. Not only is the name of the person revealed, but the reason and amount of the fine, which can be used to gather more information about him, spear phishing, and also to know if there is any activity that he has done that may be suspicious or unreliable, that can be used in any context. I remember once that an acquaintance of mine located that one of his potential clients had a series of notifications of this kind derived from multiple defaults ...



Resolución de 21 de octubre de 2022, del Banco de España, por la que se publican las sanciones de multa por la comisión de infracción muy grave a Gesvalt Sociedad de Tasación, SA, y a sus cargos de administración y dirección don Germán Casaseca Hontoria, doña Sandra Aránzazu Daza Silva y don Guillermo Masso López.

Publicado en: «BOE» núm. 261, de 31 de octubre de 2022, páginas 148713 a 148714 (2 págs.)
Sección: III. Otras disposiciones
Departamento: Banco de España
Referencia: BOE-A-2022-17817

Otros formatos: PDF XML

TEXTO ORIGINAL

Con fecha 27 de septiembre de 2022 el Consejo de Gobierno del Banco de España dictó la Resolución sancionadora por la que finalizó el expediente de referencia IE/ST-2/2021, incoado a Gesvalt Sociedad de Tasación, SA, y a sus cargos de administración y dirección, don Germán Casaseca Hontoria, doña Sandra Aránzazu Daza Silva y don Guillermo Masso López por Acuerdo de la Comisión Ejecutiva del Banco de España de fecha 25 de noviembre de 2021.

Por otra parte, el artículo 115.5 de la Ley 10/2014, de 26 de junio, de ordenación, supervisión y solvencia de entidades de crédito, establece que las sanciones por infracciones muy graves serán publicadas en el «Boletín Oficial del Estado» una vez que sean firmes en vía administrativa.

Toda vez que la citada resolución ha adquirido firmeza en vía administrativa, procede, al amparo de lo establecido en el mencionado artículo 115.5, dar publicidad en el «Boletín Oficial del Estado» a las siguientes sanciones, impuestas a los citados expedientados en dicha Resolución sancionadora del Consejo de Gobierno del Banco de España:

– Sanción de multa impuesta a Gesvalt Sociedad de Tasación, SA, por importe de ciento treinta y seis mil euros (136.000 €) –tras la aplicación de la reducción del 20 % por pago voluntario contemplada en el artículo 85 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas–, prevista en el artículo 97.1.a) de la Ley 10/2014, de 26 de

Figure 75. BOE notifying fines to people

Sometimes one can find, in the `robots.txt` files of the websites that keep these documents accessible, links to specific PDFs. This can be done so that their content is not listed in search engines such as *Google*. One of the reasons why this is happening is precisely because personal data is being disclosed, and there is legislation relating to the "right to be forgotten". However, it is still public information, so in case we suspect that something we need is there, **a script can be developed** to download those files and search for content among them if necessary. You may think that this is something quite strange, but if you think about it, it is a way of maintaining a certain "right to be forgotten" and that the information is public at the same time ...

Another thing that is also reachable via BOE or similar is when a public administration or body draws up a **list of conditions** so that other companies are able to compete and serve what that administration requires. If this specification includes serving certain **hardware / software models or types** we can then identify which machines they are using and also operating systems and software because they are linked to them. If we locate who has won the contract (it should be public information), it is a fact to add that can be used in spear phishing scams. There are several examples of CEO scams that use this kind of information...

Telephones and credit cards (or other documents of which pictures appear on the Internet)

This is a bit the same than previous sections: on the Internet it is typical to find **leaks of phone numbers** that have been used to register on a service, since information thefts from various companies or RRSS are unfortunately common. With them it is easier to be a victim of phishing or spam, etc. and they are usually

sold (or given away) in huge packs of numbers in CSV or text format, often accompanied by the name of the owner or more data. This information is not usually published in typical search engines however, but on the Dark Web.

The **credit card number** is more sensitive information than the **account number or IBAN**. But only with this information you cannot withdraw money from an ATM. You can only operate with it (e.g., make purchases online) if, **in addition, the expiration date, security number or PIN has also been obtained**. But of course, one can inadvertently leave a photo of both things on a social network by carelessness or accident ... or that someone gets the number and then the PIN/date by phishing, or simply **looking at what you type in an ATM** (hence it is important not to reveal where you usually go on a daily basis). On other occasions, people (especially young and/or inexperienced) **directly publish this data on the Internet**, and the investigated target (or one of their relatives) may be one of them. We are not talking, as before, that it appears "in the background" in a photo (which can also be), but **directly publish this information**. It is more common than you think on *Twitter*, although they are often pranksters.

Figure 76. Yes, this really happened.... and it continues to happen from time to time

What if someone has someone else's card number? The answer is **nothing**, sharing the card number does not represent any danger, since, as we said, to make an online purchase or payments it is also necessary to put the expiration date and the CVV code. The problem is, as we said, that CVV is obtained with other ways ...

If someone gets an **account number or IBAN** from a third party, they **can NOT steal from them** either (imagine then what would happen with so many donation and help campaigns that are giving that data to accept donations!). Only with that numbering no one can get the money out of your account. The problem is if **they also know the DNI number of the holde**r (in the ways we saw before). In that case there is the possibility that in some cases a direct debit of receipts can be made with this data.

Use of various means of transport

What happens if you use means of transport and publish it? Again, you are giving too much information. That also includes the **tickets** of the trip because they allow you to find out the route to follow, times, stops,

stopovers, routes ... etc. on many occasions. That is, again they allow you to see where you are going to be at any given time and, if it is a trip that you make periodically, it is even worse ...

As an example, look at this **plane ticket**. If you put it on a social network, only with the date and with the identifier of the plane (in this case AF112) there are web pages that inform you of the route, stopovers, etc. (the companies themselves have that information published of recent flights) and can even follow it in real time, such as ***FlightAware***, for example: https://es.flightaware.com/.

Figure 77. Example of a typical airline ticket

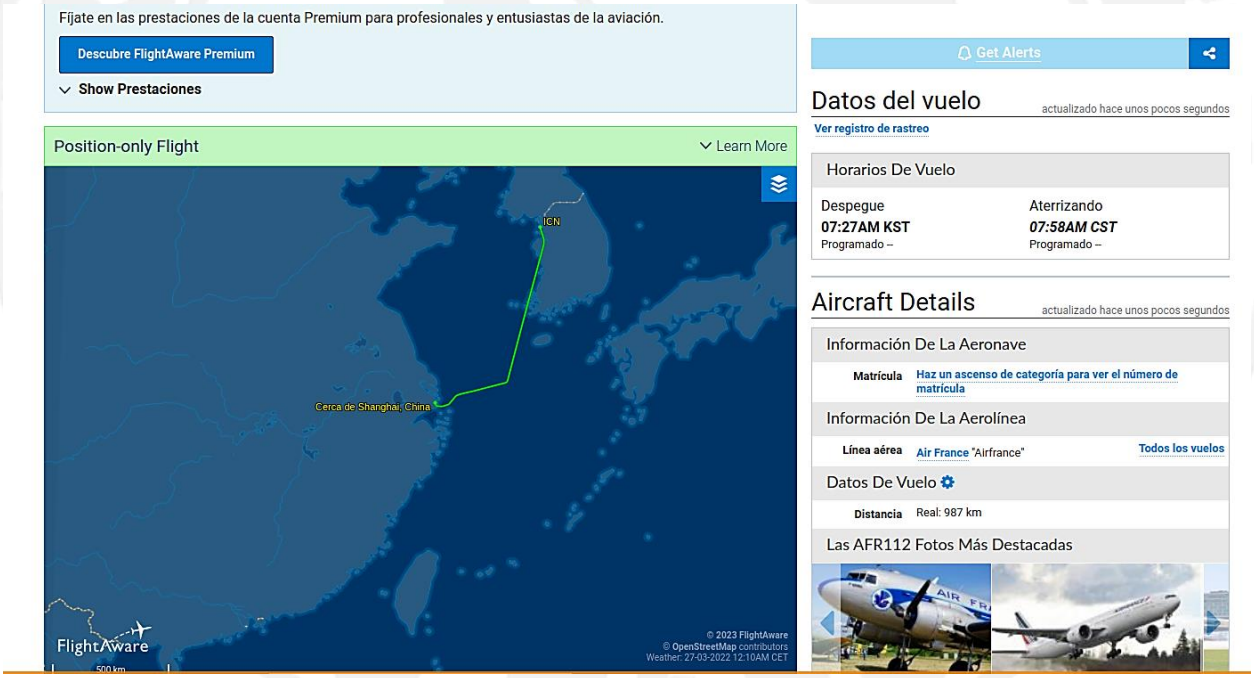


Figure 78. Flight made by the previous ticket aircraft at a later date thanks to *FlightAware*

Geolocation

In the course we have seen that it is feasible to **geolocate an IP** or coordinates that we find in metadata, but many times that is not even necessary, because the person investigated himself gives that location in the image that he uploads to the network. Do not forget this, because although it seems too obvious, people do it out of ignorance or forgetfulness. This for example is very typical in **Instagram posts**. Here there are no coordinates, they directly tell you the name of the site...

Another technique to have an estimate of the geolocation of a researcher is to use a **digital beacon** like the ones that pages like ***IPLogger*** can provide (https://iplogger.org/es/). However, this requires engagement (the objective must click on the link) and is outside the course.

Regarding IP addresses and geolocation, there are many tools to extract information associated with it. For example, ***Ghost-Track*** (https://github.com/HunxByts/GhostTrack) is a tool that allows you to obtain a geolocation of any IP from the command line **and also works with phones**.

Figure 79. There is no shortage of options to obtain geolocation information from IPs...

But we are going to stop a little more in this section to insist that many times, with a little investigative skill and knowing **common tricks in this type of research**, it is possible to find the geolocation of an image even if initially we have no clues about it. You can find the techniques used, and even a small training tutorial when applying them, in these links:

- https://www.osintcombine.com/post/image-analysis-and-verification
- https://infosecwriteups.com/image-geospatial-osint-7cf8560b54f6
- https://medium.com/@ashfquejahan5/image-analysis-osint-tools-2bad341a3d29

With techniques like this, you can find information even in frames from famous series (when instead of photos you have videos, frames are analyzed individually). For example:

Figure 80. Where *Big Bang Theory* takes place using just one frame from the series

The boom in **artificial intelligence** is also having its influence on the OSINT world, and proof of this are geolocation AIs such as ***Geospy AI.*** https://geospy.web.app/. This AI analyzes vegetation and other elements of a photo to tell you where it thinks the photo was taken, giving you coordinates that you can use on *Google Maps* to validate the information. It is a beta project, and **its results are approximate**, but to get a first clue of the area where a photo could have been taken, it fulfills its function perfectly (in addition to being free 😊). Here is an example with a photo I found on social media where you can see a minimal background:

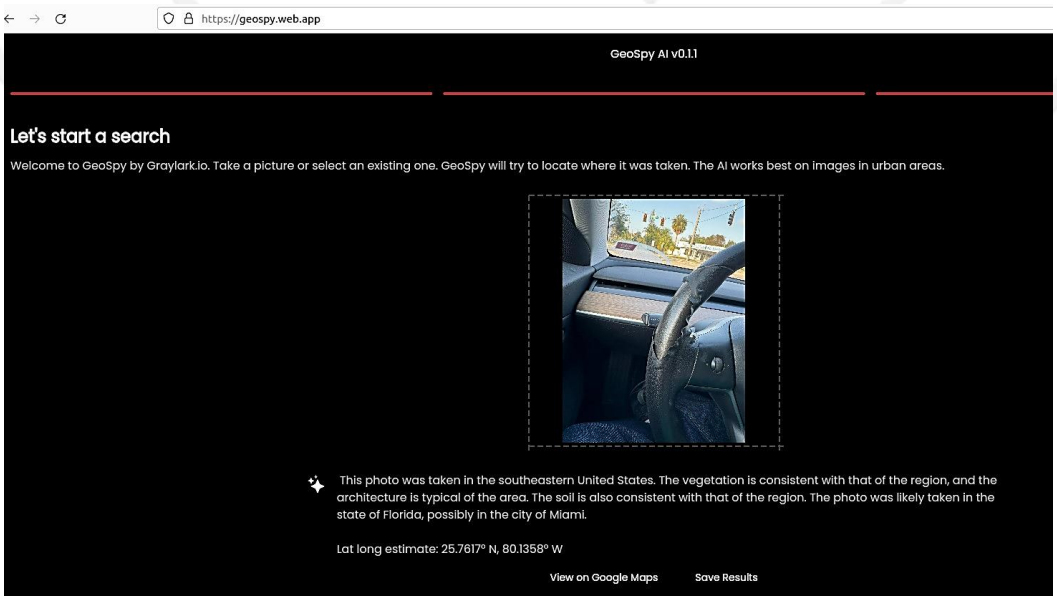


Figure 81. I admit that it does a lot with as little as the photo shows...

And this is a photo of Adrián Barbón in an Asturian town. It is clear that he did not get it right, but at least we know that it was taken in the north of Spain (keep in mind that if someone from abroad sees this photo they would not even know where to start, as they would not even know who the one in the photo):

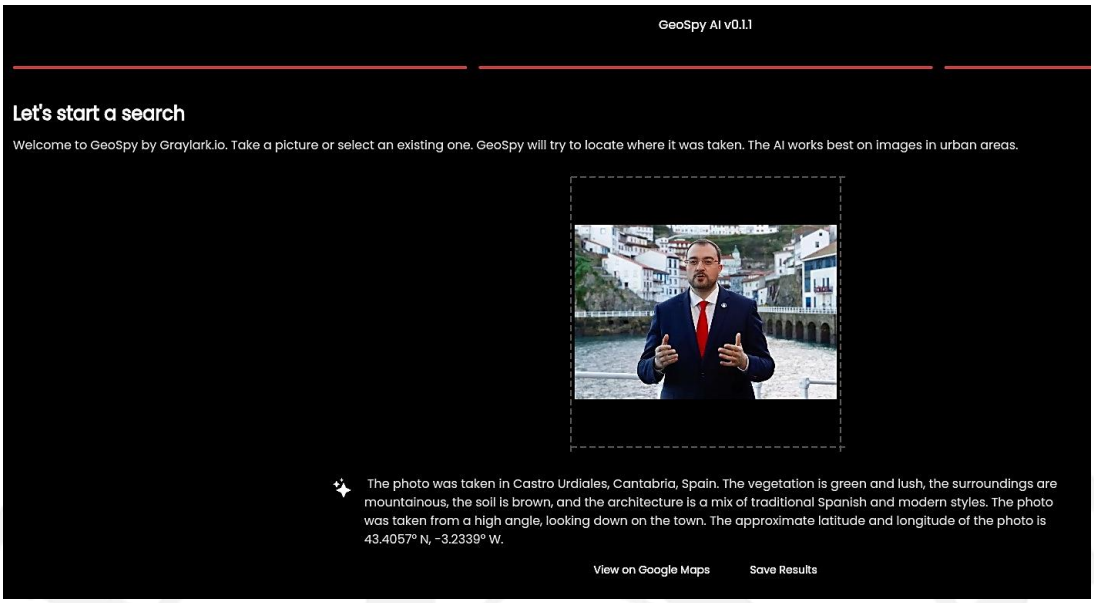


Figure 82. The results are approximate... But at least now you know where to start looking

Forums / Blogs / IRC...

As we said at the beginning, our targets can write relevant or interesting data for certain research in forums, blogs, IRC channels, *Discord*, *YouTube* comments, etc. It is common for a specific person to use the same ID in many social networks or pages, so locating it is viable at the time we use tools such as ***Project Sherlock*** (**Seminar 1**) or because simply that information is being given on other websites without realizing it.

We must also bear in mind that many people **"lose inhibitions" on sites that are not social networks** (that is, they think or say things that they would not dare to say in more public forums, demonstrating a tremendous hypocrisy), protected by that **supposed anonymity** they have in those sites, thinking that nobody will be able to find out their real identity. It is common, both by criminals and by state security forces, to use a false identity, like the ones we saw before, to become a user on these sites and "cyber-infiltrate" simply to read what is commented on in those forums in search of valid information for something.

In addition to the **coincidence of IDs** used in different places and / or possible metadata that there is in things that they upload, it is also sometimes feasible **to identify them by what they write and how they write it**: specific expressions that are typical of the person, "rare" spelling mistakes that the person usually commits when writing, use of particular words, narrate the same personal story (or copy paragraphs) in different places, ... An analysis of what a person writes can reveal more than you think... Moral: Never underestimate the power of a **forensic linguist**, such as Sheila Queralt: https://es.wikipedia.org/wiki/Sheila_Queralt_Estevez.

As an illustrative anecdote of this type of technique, we once identified a troll on the Internet because, to try to remove suspicions from his real identity, one day he decided to criticize a specific and identifiable person in his social networks. The failure was that of all the people he criticized from his "anonymity", this person was an "**outlier**", that is, he had no traits in common (for example, positions or job category) with the others he criticized. With that information, and others that he revealed as he unfolded his trolling campaigns (always following the 4 commandments of the OSINT that we saw before), we were able to make

a profile that reduced the number of "suspects" to 1. Brought all this to the attention of the authorities, it finally turned out that he was the real culprit and the campaign stopped. Indeed, in the end the culprit was the "outlier" whom he had criticized very harshly in networks (that is, who criticized himself 😊). Maybe criticizing yourself to redirect suspicions may seem like a strange tactic, but it is not as uncommon as you might think...

Announcements

Searching for a person on an **ad or sales portal** (through photos of properties or cars for example that they have uploaded to the network, or a common user ID with their social networks) is an alternative way to locate and obtain additional information of the target. We must also include in this assumption **companies that are looking for employees of a certain profile** or with knowledge of some technologies, since that can be used to **identify software or hardware** that the company is using at that time. If the company, for example, asks for many people with a certain profile, it will mean either that it is going to open a new line of business, or that the people who were occupying it have been fired for some reason, which opens the possibility of looking for former employees to continue with our investigation. Another case that I have seen in my experience are **companies that ask for workers of a certain profile too frequently**. This often reveals a bad or toxic work environment, and that generates a large number of "burned" ex-workers, with all that this implies.

Therefore, the information that can be found in second-hand sales portals, properties, real estate, employment or similar should not be underestimated because, if we can find a link with the person we are investigating, we can even **open a communication channel** under a false identity such as those mentioned above to extract more information from it. However, this is a technique that requires "engagement" and therefore is outside the objectives of this course. Again, it is only mentioned just in case one day you are the target. **Scams on sales pages are all too common...**

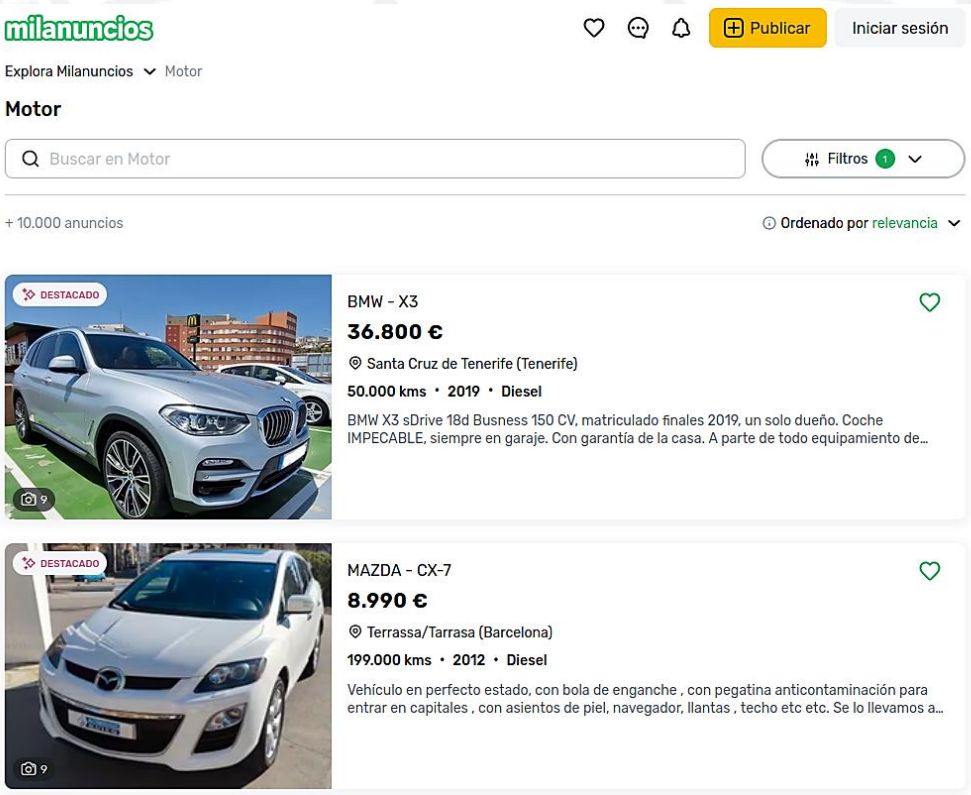


Figure 83. Never underestimate the information that can be found on a second-hand sales portal or similar...

HACK FIRST • HACK HARD
大和
Cobra Kali
• NO MERCY •

OSINT on machines and code

I can tell you little new here, since it is based on applying the techniques that were seen in **Seminar 1** of the course and were also tested in Labs 1, 2, and 8-9. *Shodan*, *Censys* and *Google Dorking* on **machines** are typical passive ways to do this and `nmap` the preferred active form in these cases, both with standard scans and through scripts. For **code** on *GitHub*, **Seminar 1** reviewed tools and techniques.

It is recommended to look at these topics of the course to do this type of research, **always prioritizing doing it passively**. You should also consult the section of this document where search engines that offer various information, which may be applicable in this case, are described. If what we saw from the GHDB seems complex **to you, you can use an AI that works with this concept** (make *Google* queries that find interesting information about objectives) and/or you want to propose some personalized ones, you can use *DorkGPT* (https://www.dorkgpt.com/), which says what you want to find generates queries that search for it.

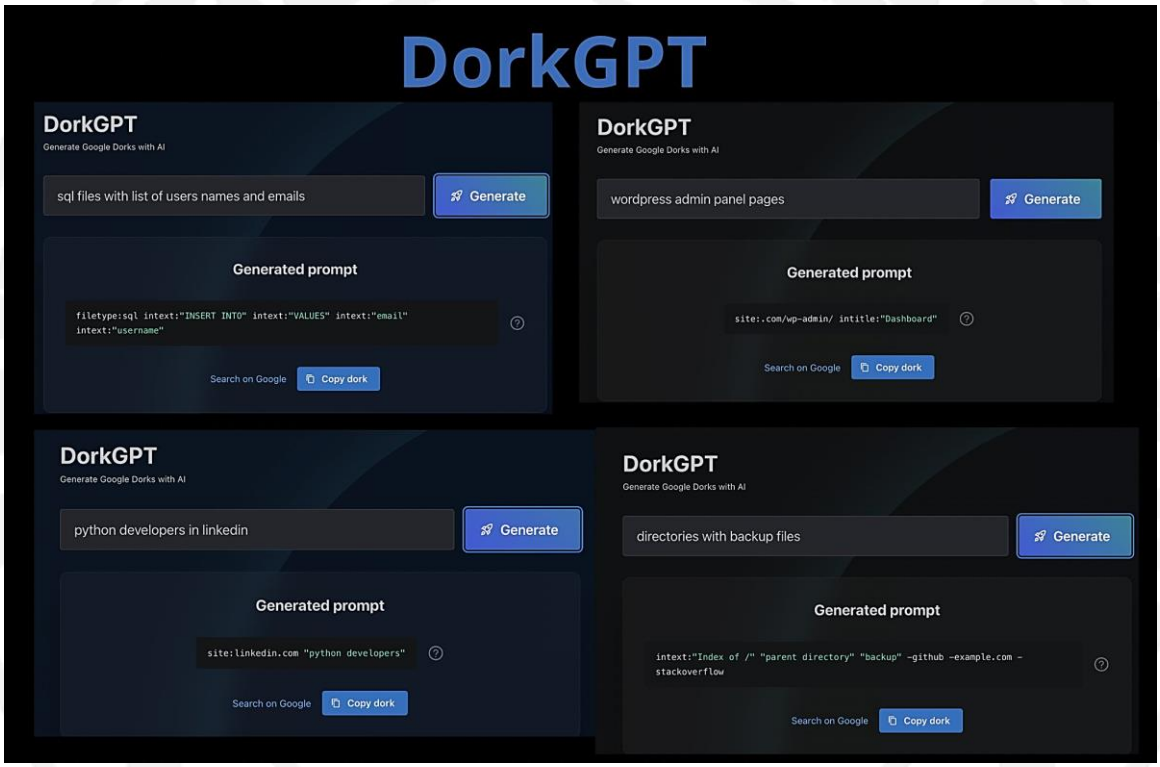


Figure 84 *DorkGPT*, using AI with *Google Hacking*. Source: https://twitter.com/cyb_detective/status/1654835052260335618

Obviously, if the information of machines and code is obtained from, or linked to, what we have obtained by **examining the information of people**, the study will have a much better chance of success (and quality). On this particular section we will highlight only a few things in the following subsections.

Mobile terminal emulation

You may think that it makes no sense to access a website as if you were on a mobile for OSINT research topics. But trust me: **it is common that a mobile website does not have the same protections or security mechanisms as the ones targeted for PCs**. In addition, it is also possible that the web for mobile devices is independent of the other, and it may have vulnerabilities or extra information (such as documents) that the other does not have. **It is not necessary to use a mobile to see these versions of a website**, there are browser extensions (or the functionalities integrated into them) that allow you to see them from a PC:

https://stackoverflow.com/questions/16651911/how-can-i-simulate-mobile-devices-and-debug-in-firefox-browser

IP Addresses / Domains

We already know that, with the IP of a target, we have a way to easily find out the services it offers and its version (`nmap`, **Laboratory 7-8**) and, therefore, whether or not they are vulnerable (**Topic 1**). The search engines seen in the corresponding section of this document can also be very helpful in finding more information about the targets whose IPs we know or find out, as well as the ***IPInfo*** website (https://ipinfo.io/). As we can see, the IP of a target can become a "mine" of information.

Other things we can look at from IPs / domains are:

- If they appear on **blocklists**, which would mean that the target could have been compromised or their activities are not very "normal". There are many block lists available on the Internet, downloadable or searchable online. A very popular and large one, classified by the type of activity carried out by each IP, is maintained by Steven Black https://github.com/StevenBlack/hosts
- **If they are protected by a CDN** (e.g., *CloudFlare*, *TransparentEdge*), which implies that there is an investment behind it, so security is taken “somewhat” seriously.
- If they belong to a large or small **cloud/hosting provider**. This could give them greater initial security, but also **volatility**. Volatility implies that they are domains that are created only during small periods of time, usually because they are for scams or similar. You buy the domain (usually very cheap), give it a name that "sounds like" the original (or send victims a shortened link), clone a page to impersonate, launch the campaign and in a few days delete everything to hinder investigations (***TA0042 Resource Development*** of MITRE ATT&CK). If in your investigations you find a machine in the cloud of, for example, ***Amazon***, with a copy of a known page, you have found a scam.

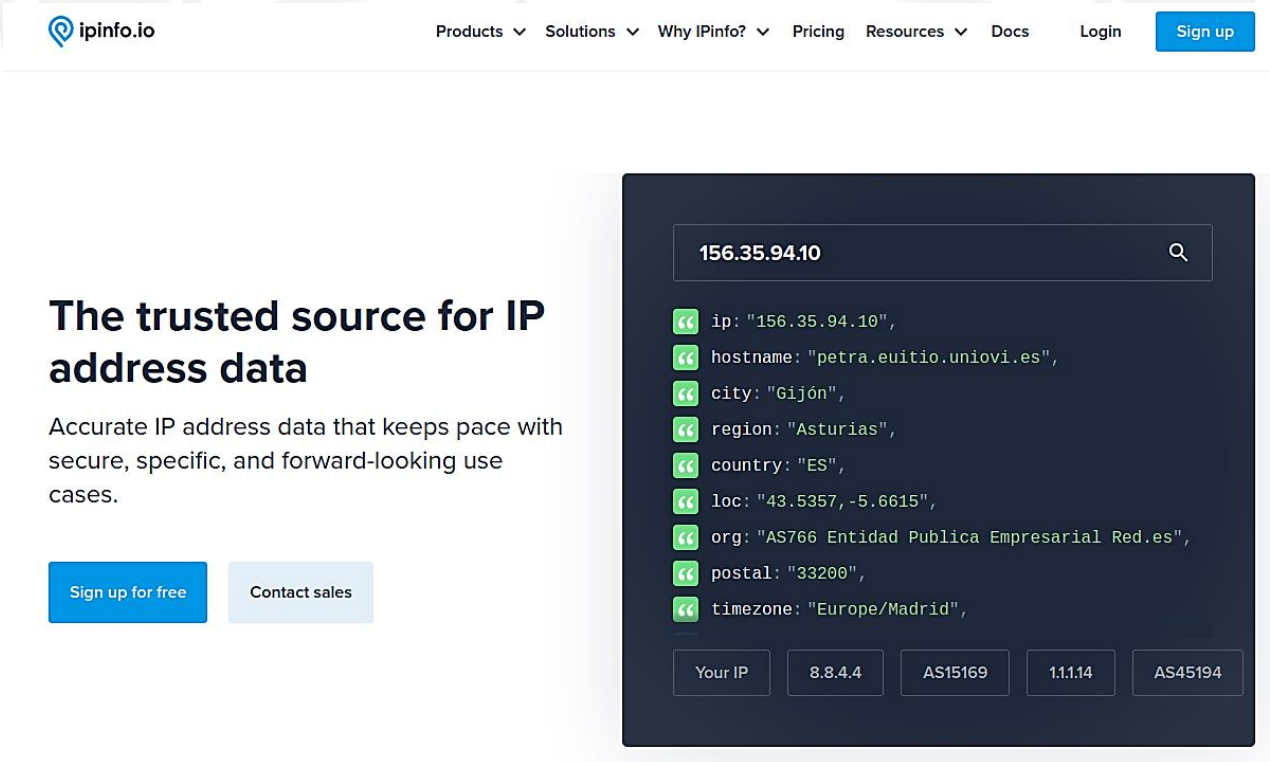


Figure 85. IPInfo Web

If you want to **know more about a domain and the "secondary" sites that it accesses or read information from** without having to visit it in case it had some kind of malware, you can always resort to ***Lookyloo's*** services: https://lookyloo.circl.lu/

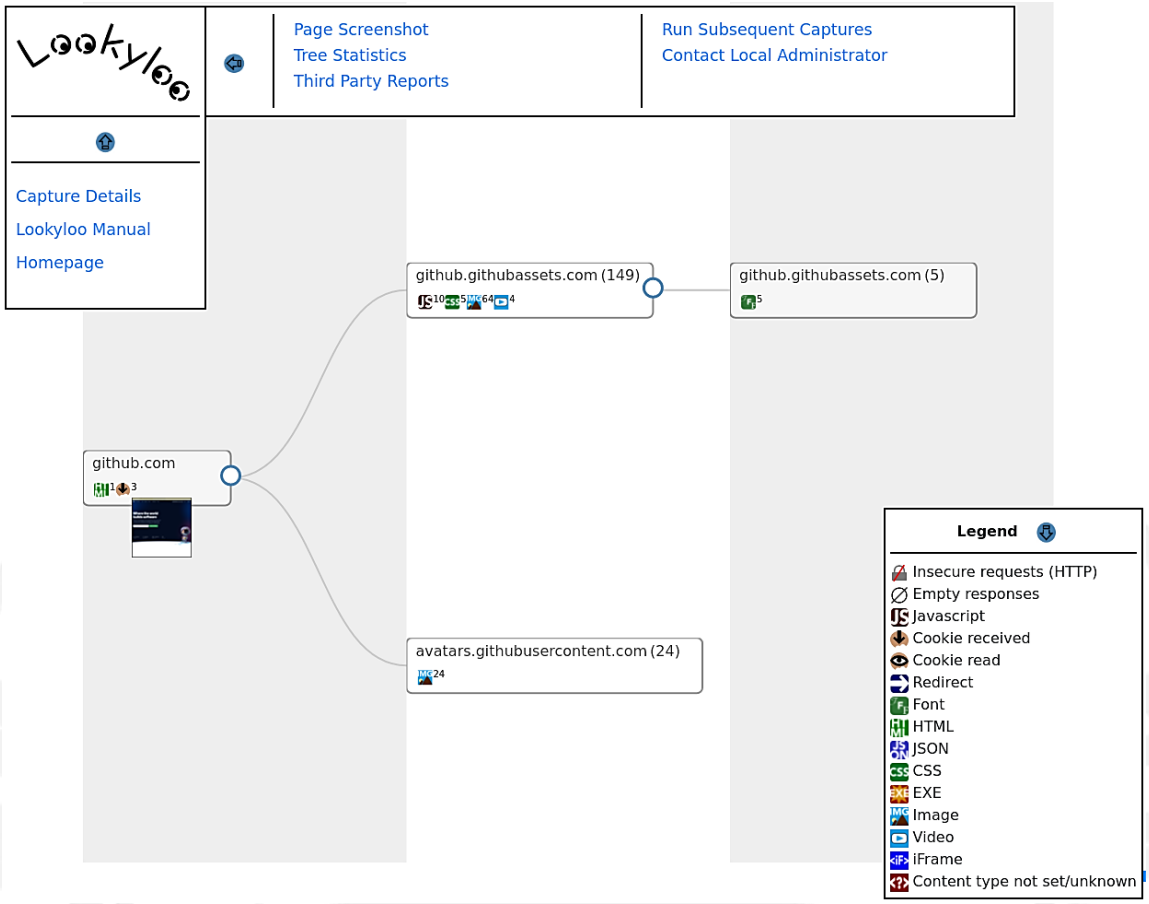


Figure 86. *Lookyloo* analyzing what a domain does. Source: https://github.com/Lookyloo/lookyloo

Offline analysis of files on the network

Examining the data of the files contained in a page is obviously a source of information that we have already seen in the course. If we include it here it is simply to remember that many times it is an investigation that takes **time and several work sessions**, and in those cases it is best **to make a local copy of a website** to look at what is published calmly. This is possible thanks to tools such as ***Burp Suite*** (installed by default on ***Kali Linux***), that allow you to **make local "mirrors"** of any website with a single click: https://portswigger.net/burp/documentation/desktop/getting-started/running-your-first-scan

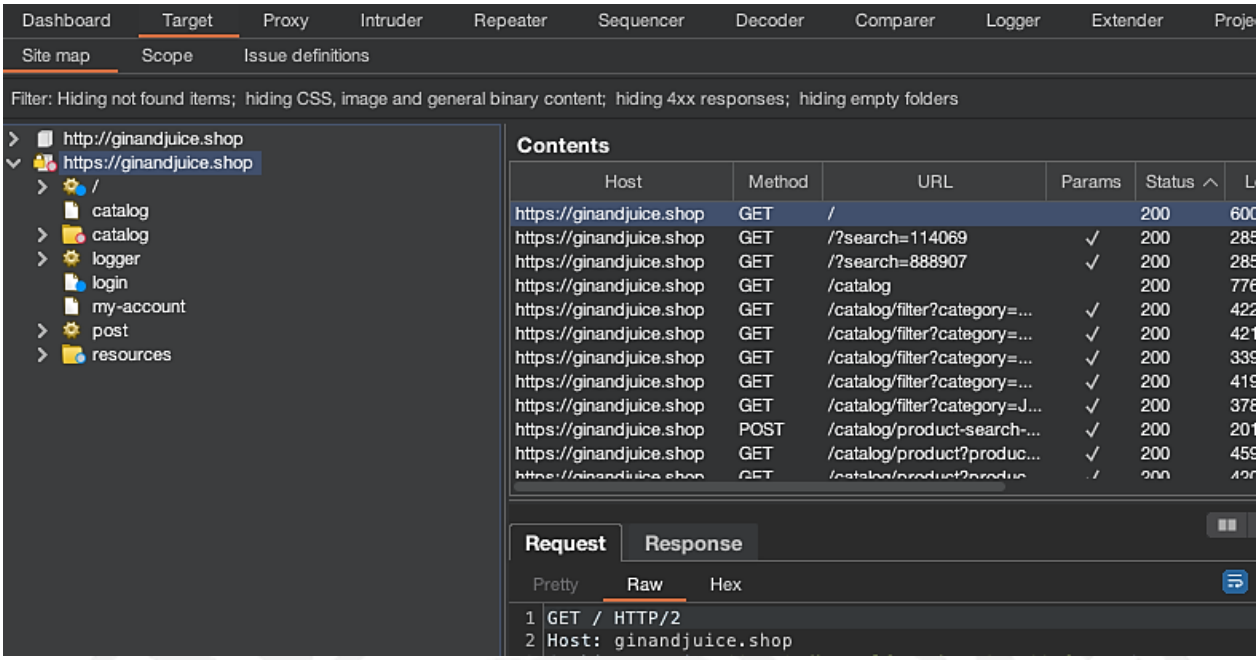


Figure 87. Burp suite making an offline copy of a website

With the website in our possession, we can do all the searches we want at home without haste and at our own pace 😊.

In addition to data that can be found in comments on the downloaded website, it is possible to find other more subtle information if we know exactly what to look at. For example, in the following image we can see an example analyzing the code that the website has associated with ***Google Analytics*** (a tool used to analyze the behavior of users on the web). With the *NerdyData* (https://www.nerdydata.com/) tool mentioned in this image you can find other websites linked to the one you are visiting... And those other websites may have more information about their owner, associated companies, etc.

In other words, we are using *Google*'s analytics identifier to find **websites linked to each other**. *Why will they be linked?* It is up to you to know . In addition, you now have another tool (*NerdyData*, which can be tried for free by signing up for an account) to **match websites that use the same SaaS technologies**.

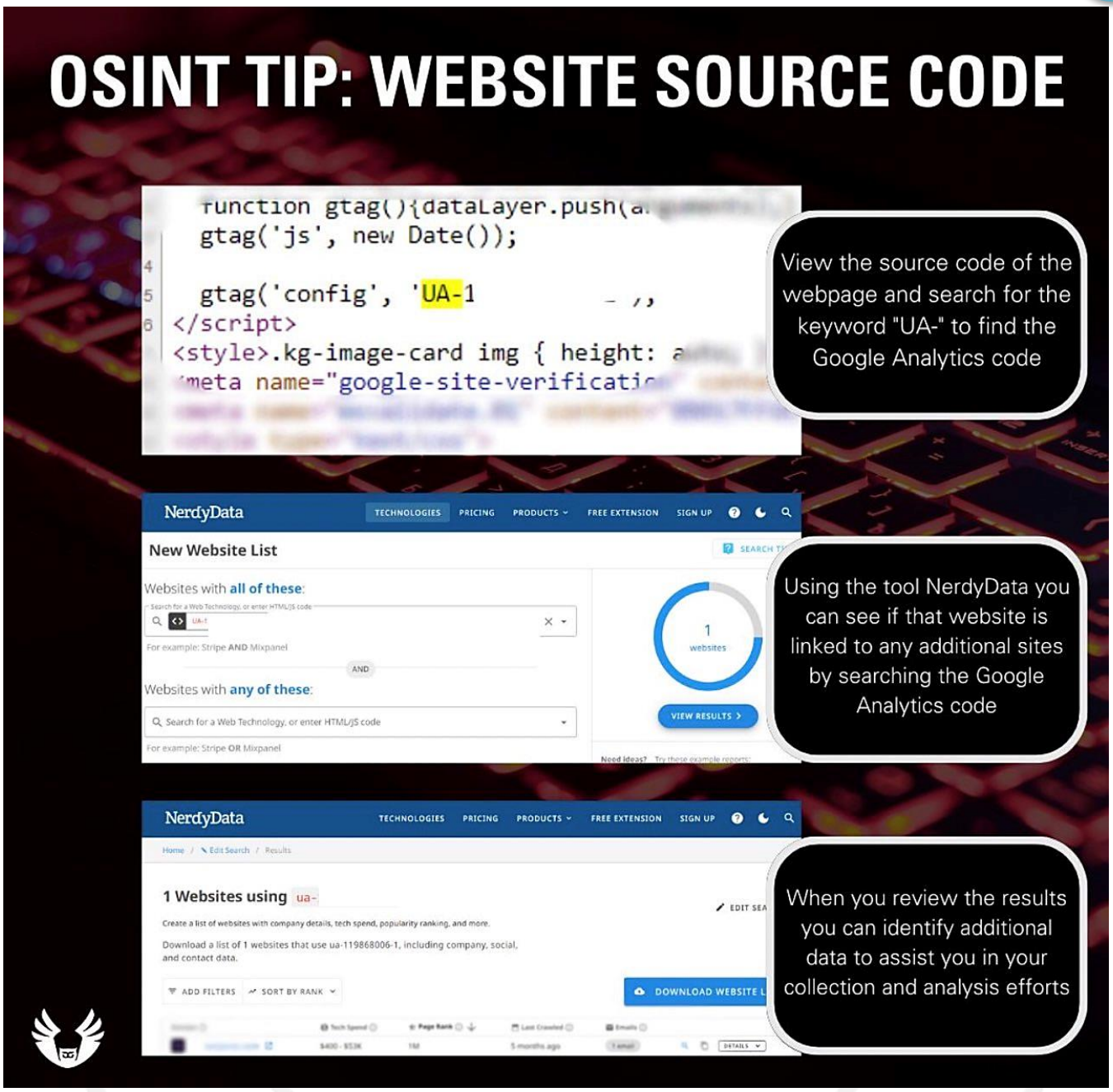


Figure 88 . A website's code can be an unexpected source of data... sometimes surprisingly and without requiring negligence on the part of the perpetrators. Source: https://twitter.com/cybernitewatch/status/1664975350999252992

What happens if you leave your devices unlocked

Everyone imagines that *if malware* is able to enter our system we have a serious problem, although many people do not have a clear idea of **exactly what can be stolen from there**. The reason is that they do not know exactly what kind of information any of our devices have, and this is clear in the image below.

However, before showing it, it is important to remember one thing: the same type of access is possible **if a work device is not blocked** when we are absent from it, and it will not be the first or the last time that someone takes advantage of that to get information from someone. This is not OSINT, nor something we are going to do in the subject, but it is important to know it because it is a very real risk and often quite ignored (in other words, **you can be a victim**).

The following image applies to citizens of the United States, but many of the elements in it are applicable to everyone. Think about the number of remembered passwords you may have, your history, personal documents, etc. If it is already bad to have it stolen using malware, do not allow them to do it by an oversight: **it is not a joke**.

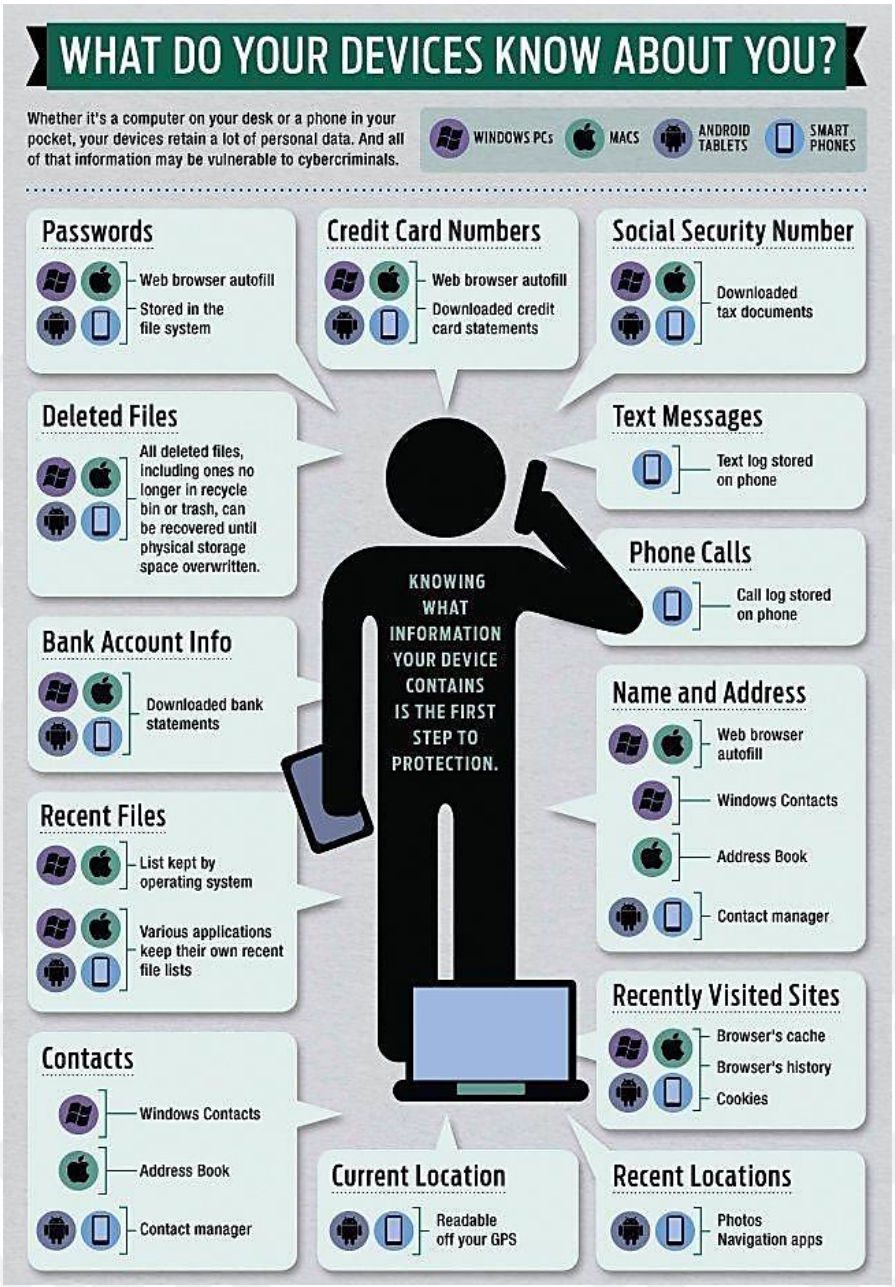


Figure 89. Things that can be taken from you if somebody has access to your devices. Source: *Cyber Crime watch*

"Automatic" OSINT. OSINT Ex Machina

OSINT on people

Before we mentioned social networks, but the analysis that can be made about what they contain **can be very broad and revealing**, greater the more things the investigated person shares through them. That is why it is appropriate to complement this analysis with automatic tools. To do this we must use **specific tools for the concrete social network** that we are investigating, such as those seen in **Seminar 1** and those listed on the ***Ciberpatrulla*** website (https://ciberpatrulla.com/), which allow us to take a historical list of uploaded information, geolocations, search among the written text and other data. They allow us therefore a more in-depth analysis of what appears in each social network, it can also reveal many things.

However, it should be clarified that although **there are automatic tools for all social networks**, many of them depend on being able to access to the API of each of them, which forces us to create an account in each social network and may have possible usage / ratio limitations that may interfere with our operations. In addition, as with the case of ***Twitter***, access to the API **can be not free**, and prevent its use completely if we do not have funding to do it. In this specific case, we need tools that do not use the API to extract the information but are ***scrappers***: they navigate the web interface of the application in an automated way, and extract information by reading the contents of these web pages, structuring the information properly so that we can then use it with other programs or analyze it more closely. A good example of a *scrapper* tool that works with various social networks is `snscrape`: https://github.com/JustAnotherArchivist/snscrape

As a sample of tools that work with people, the ***Email Info*** (https://epieos.com/) tool can help us locate information associated with an email we give, which can also be used to start more in-depth investigations.

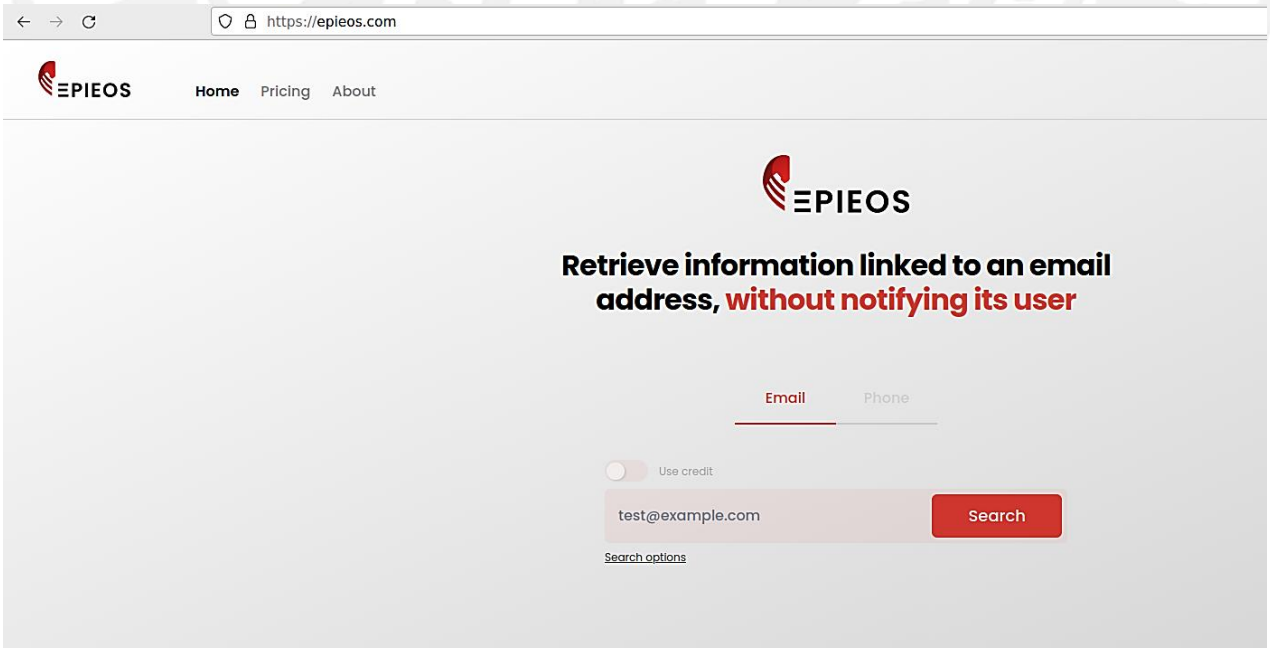


Figure 90. *Email Info* tool

On the other hand, we have ***WhatsMyName*** (https://whatsmyname.app/), which does the same but with usernames.

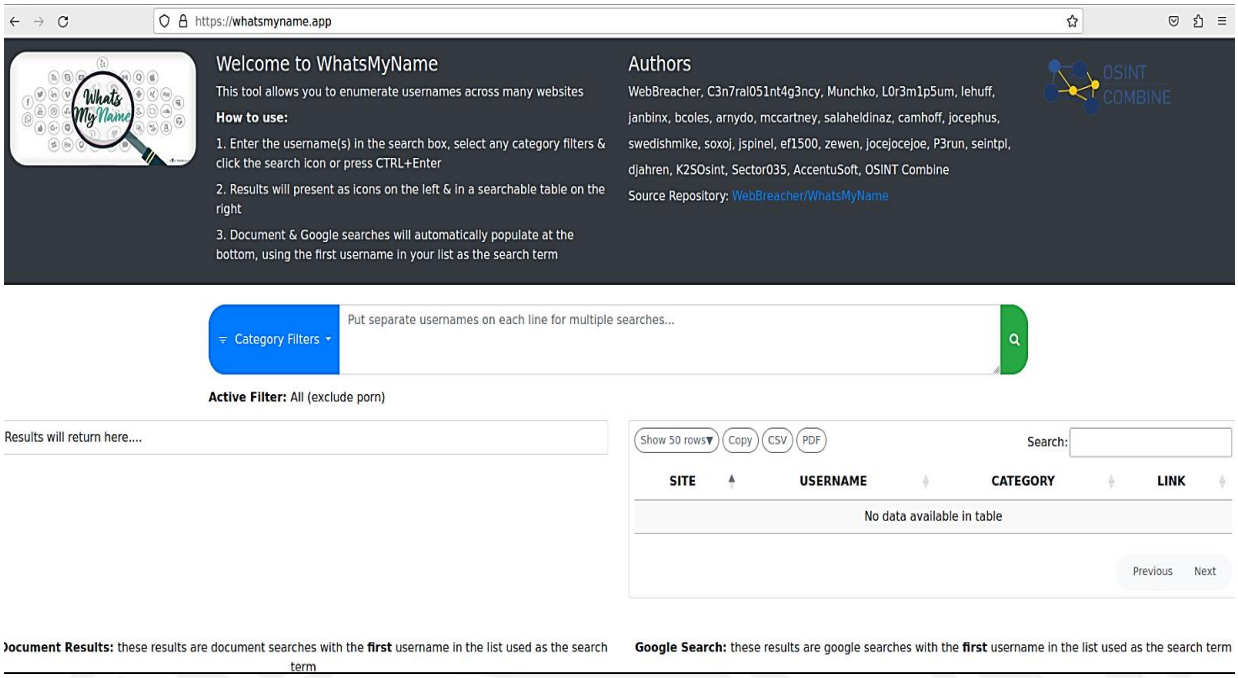


Figure 91. *What is my name* tool

Finally, there is an automatic way to find out data from a **person abusing the contact functionality** of our mobile phone, in combination with applications such as *Snapchat*, *WhatsApp*, *Threema*, *Strava*, *FitBit*, *Duolingo*, *TikTok,* or *Vivino*. It consists of:

- If you have a phone or email from the person you want to investigate, **add them as a contact** on your phone
- Use one of the apps mentioned above and add this new contact to that app if it does not already do so automatically.
- The app will show all the information of that contact, which often includes a photo, full name, and location.

These techniques and more automatic techniques to investigate people you can see documented **in this video,** which includes **obtaining information from the *dark web***: https://www.youtube.com/watch?v=ywUxkvCK96w. Although this may seem to have a single criminal purpose, it is not: many organizations do it because they know when they are **leaking data stolen from themselves**.

OSINT on machines: Mass metadata with FOCA

Introduction

Given the enormous importance of the information that can be obtained with metadata from the resources of a website, **we are going to use a tool that allows us to automatically obtain and relate the metadata of an entire website**. ***FOCA*** (*Fingerprinting Organizations with Collected Archives*) is a tool used to find metadata and hidden information in the documents it examines, which are usually found on web pages and which FOCA can download and analyze. It allows you to analyze many types of documents, commonly *Office*, *Open Office*, or *PDF* files, although it also analyzes *Adobe InDesign* files, or *SVG* for example.

These documents are searched using search engines such as *Google*, *Bing*, *Shodan* and *DuckDuckGo*. The integration of the information found by these search engines makes it possible to obtain a lot of data. There is also the possibility of adding local files to **extract EXIF information from graphic files** and make a complete analysis of the information discovered from a URL. **FOCA unifies the metadata information extracted from all the files**, trying to recognize which documents have been created from the same machine, which servers and clients host them, the people who have edited the documents, etc. In other words: as you study the documents on a web page, you make a "map" of everything you can find out from them.

What is the difference between this and what we have done in the laboratories with `exiftool`*?* Precisely, that it can automatically relate the metadata found in each and every one of the documents on the web. Therefore, if you come across many documents with metadata like the ones in this example:

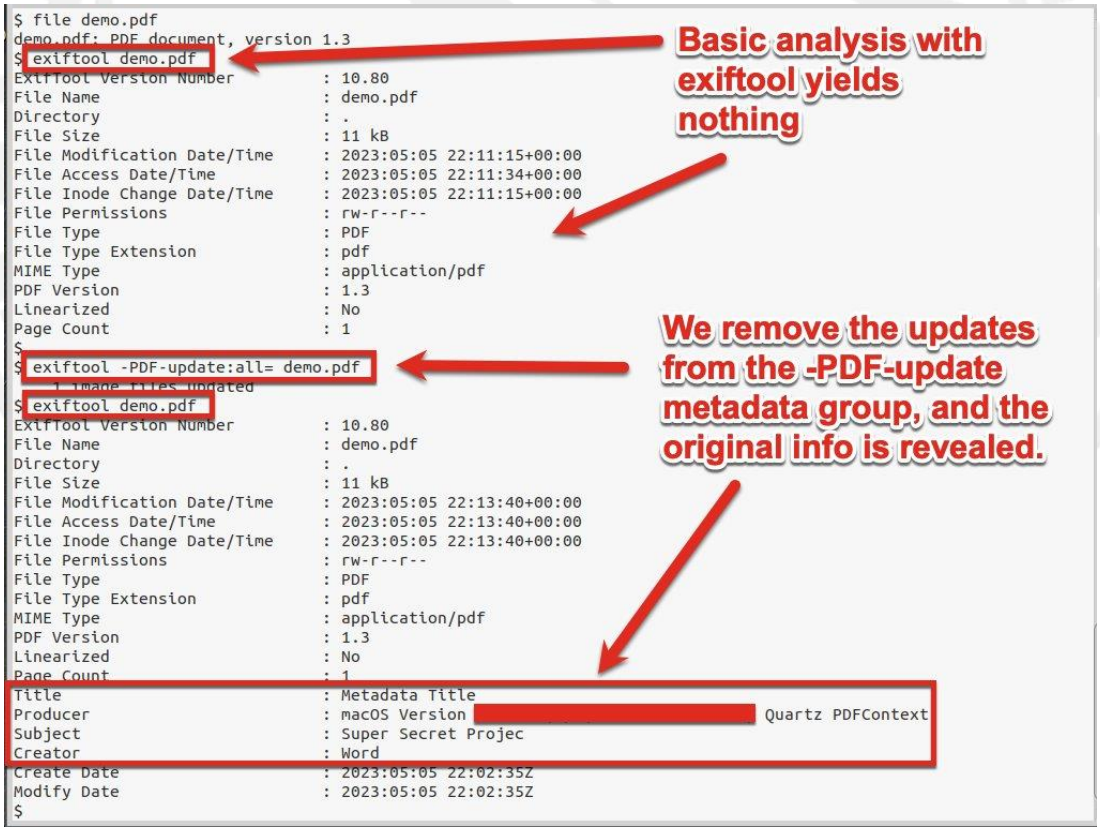


Figure 92. Extracting metadata from an individual image in an advanced way. Source: https://twitter.com/cyb_detective/status/1705261248512712781

The amount of information we could find can be enormous.

Installation

To use FOCA, **we need to use a virtual machine or Windows 10** system to install it. In addition, it is necessary to have a working installation of *SQL Server* before installing it. The evaluation *SQL Server* downloadable for free from here (option "on premises") works: https://www.microsoft.com/es-es/sql-server/sql-server-downloads

Once this is done, we can proceed to download, decompress, and run the installer from the official repository of the tool: https://github.com/ElevenPaths/FOCA. On first boot, the tool will ask for credentials to access the installed database. The correct ones are:

- ***Server name**: localhost*
- ***Authentication**: Windows Authentication*

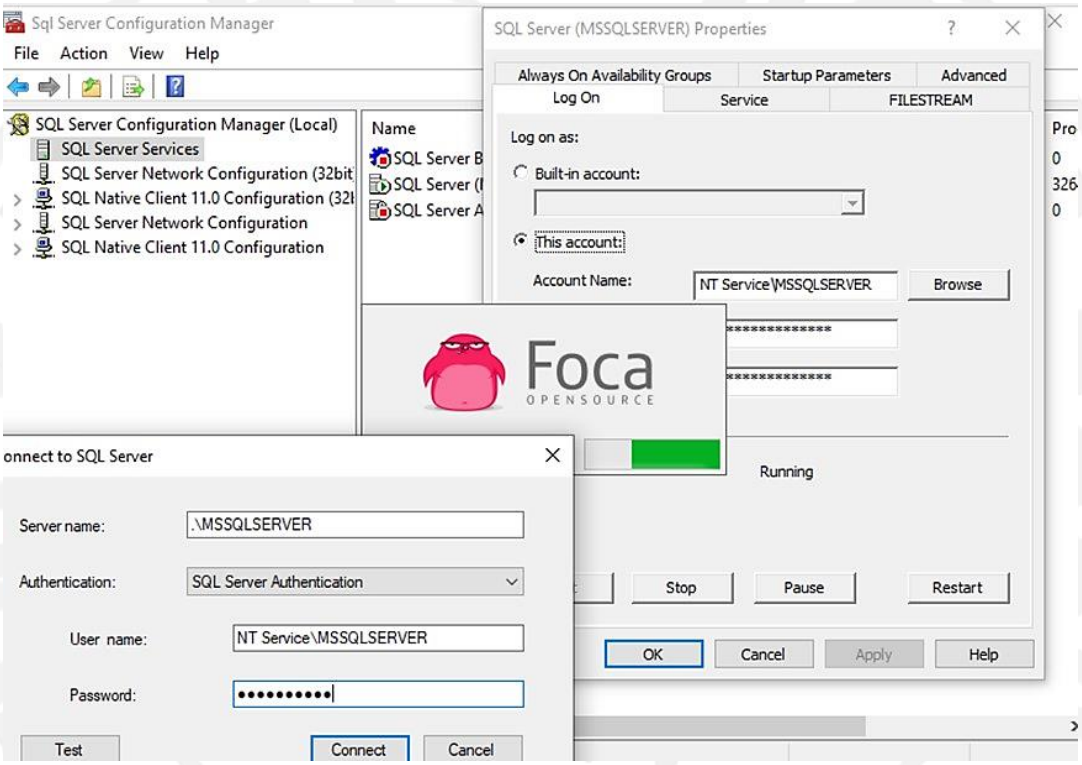


Figure 93. Initial start-up of FOCA

With this we could already enter the application and start using it.

Using FOCA

The first thing we should know is that FOCA works based on **projects**, each of which we will associate with a website that we want to analyze. To do this, at first we created one as follows.

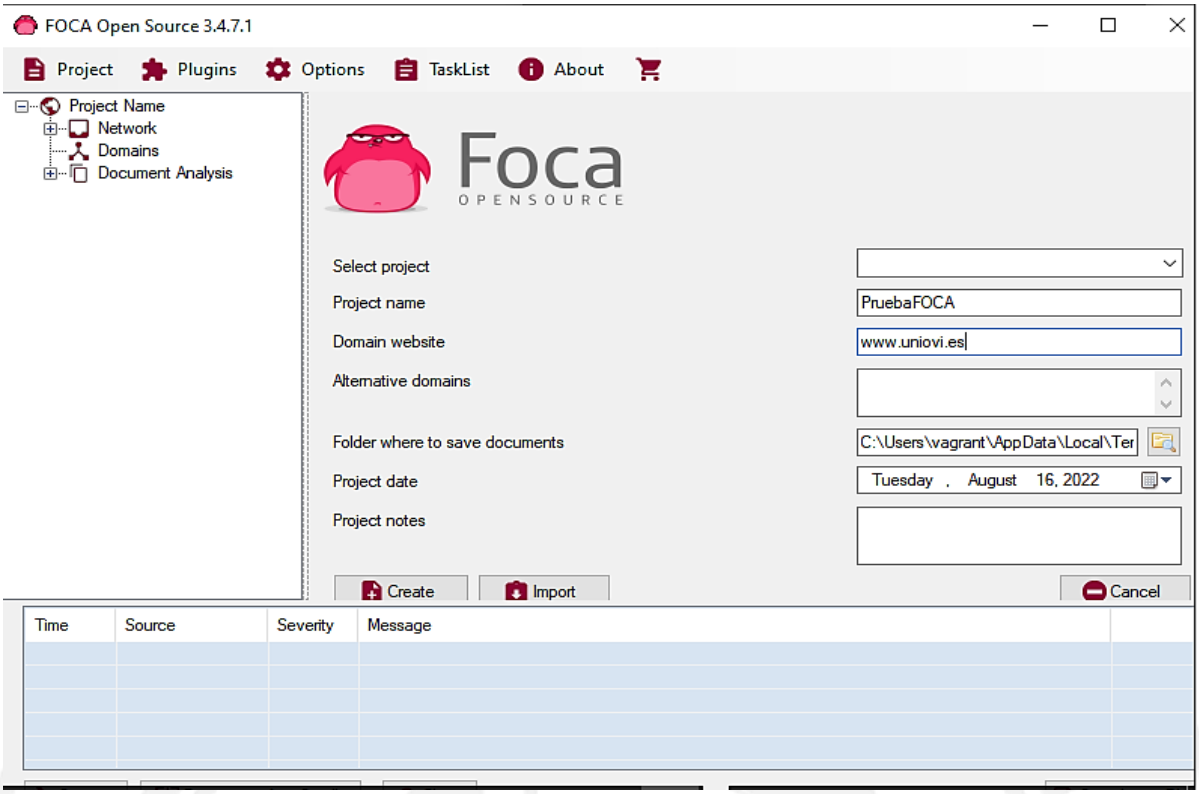


Figure 94. New FOCA project

Once this is done, we can use **the Search option** of the tool to analyze the documents on the web.

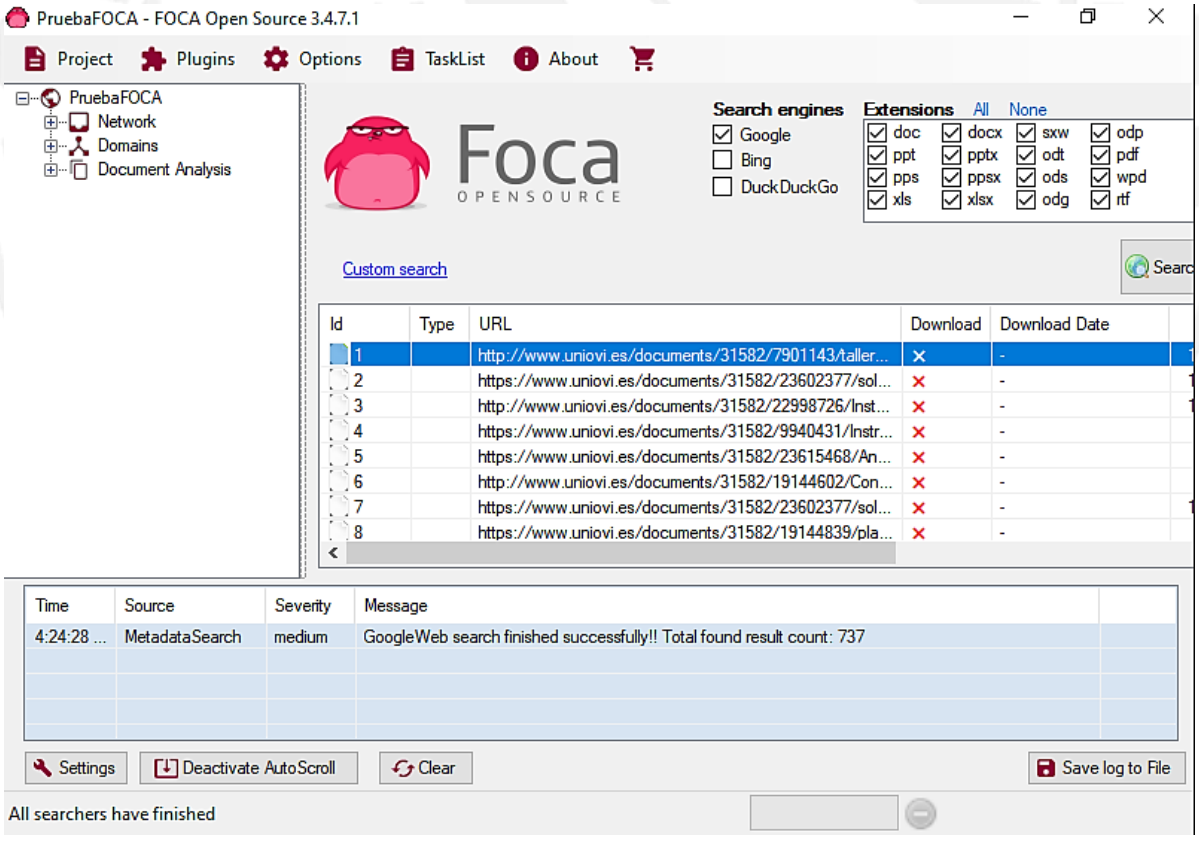


Figure 95. Searching for files on a website with FOCA

Each document obtained can be easily analyzed **with its context menu with the right button**, for which we can first **download it** (*Download*) and then **extract its metadata** (*Extract Metadata*). Note that just below **we**

have the option to do the same with all the documents found on the web, which is appropriate if we do not want to make a detailed analysis of each one.

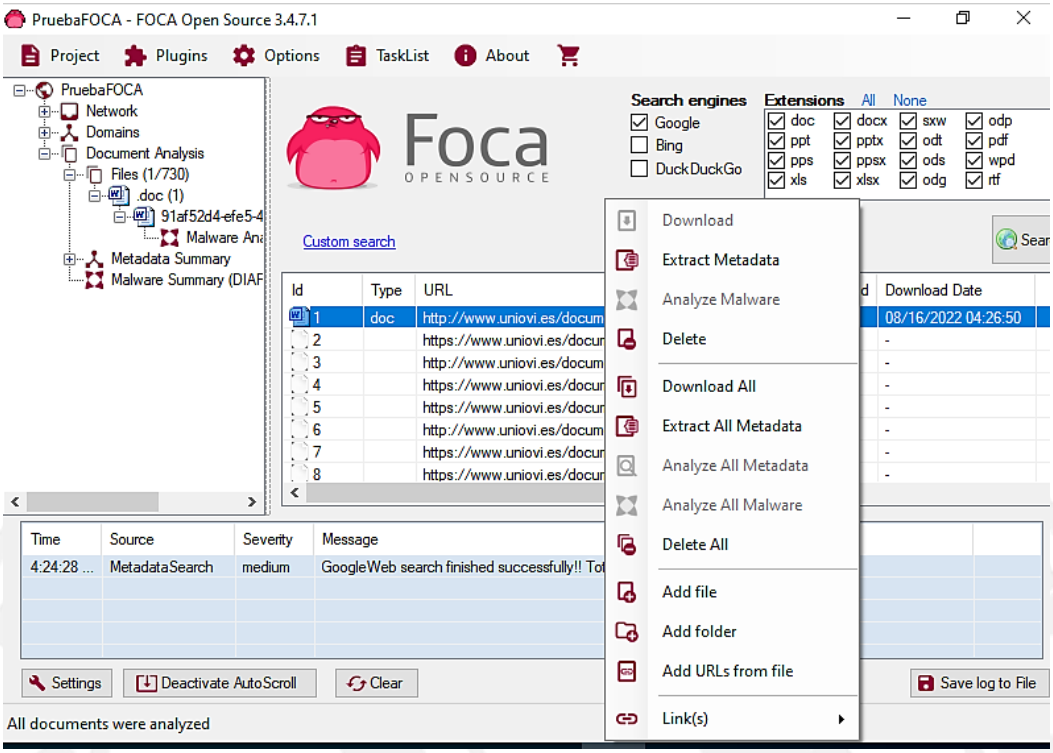


Figure 96. Extracting metadata with FOCA

When extracting the metadata, we see how now it gives us all the information that has been found in it that we have mentioned above.

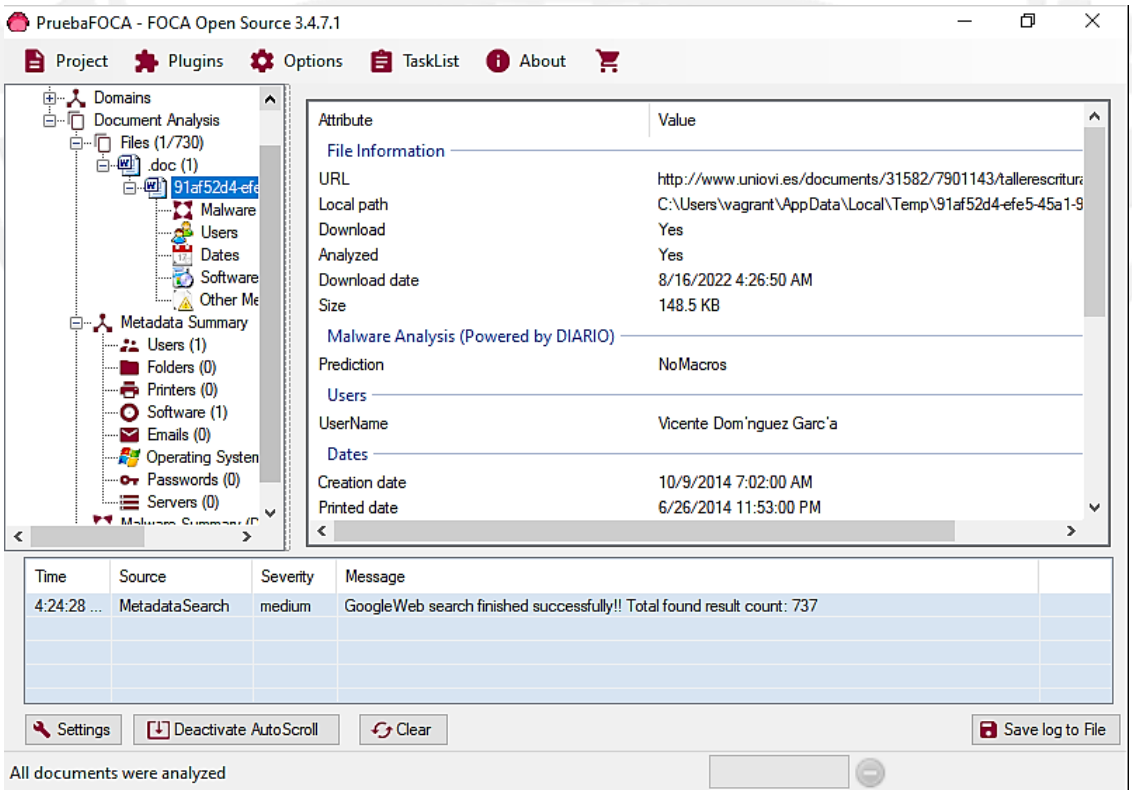


Figure 97. File metadata extracted with FOCA

One of the most interesting aspects is that **it has an integrated antimalware system** to be able to check the possible danger of each file. We also see sections for other common metadata such as users, dates, software, etc.

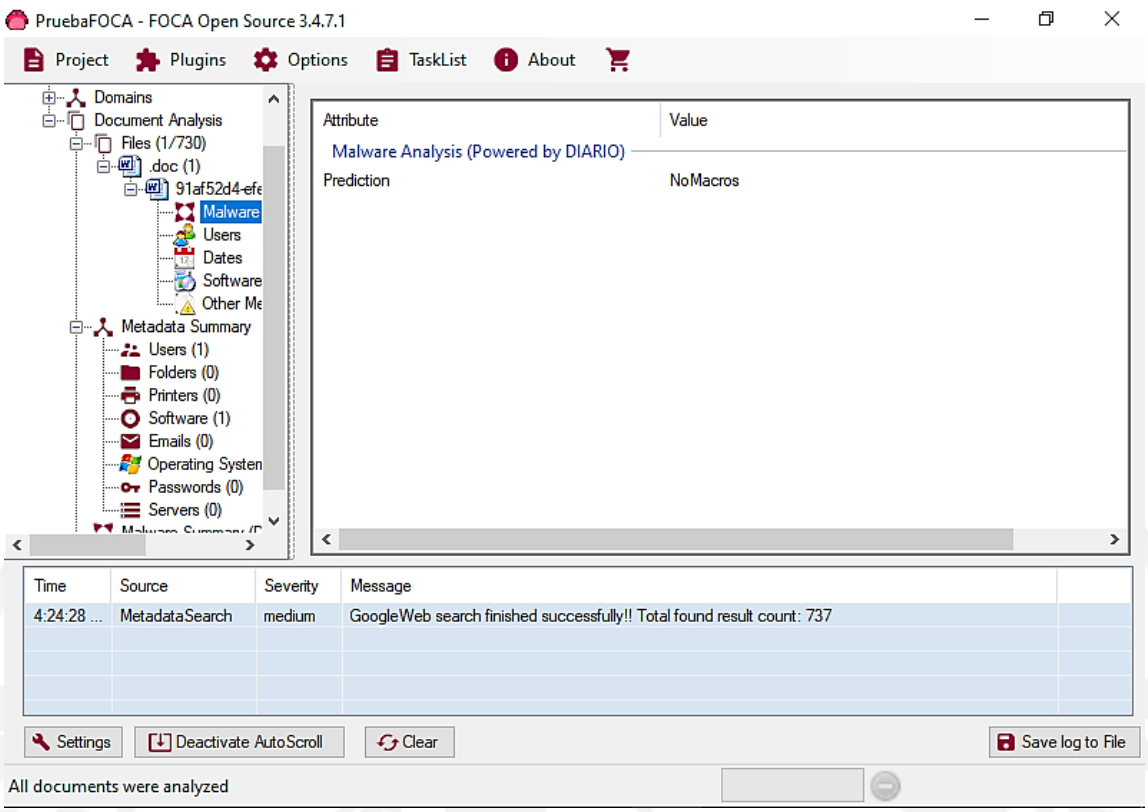


Figure 98. FOCA's antimalware

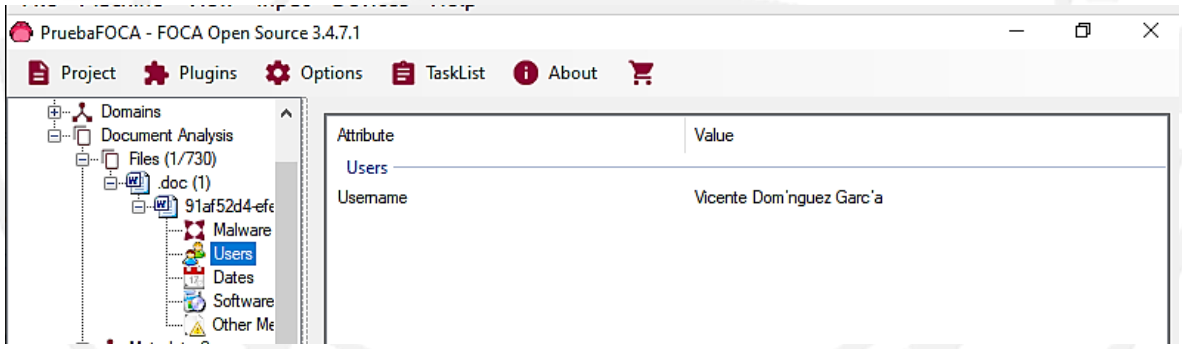


Figure 99. FOCA metadata by groups (I). Users that created the documents

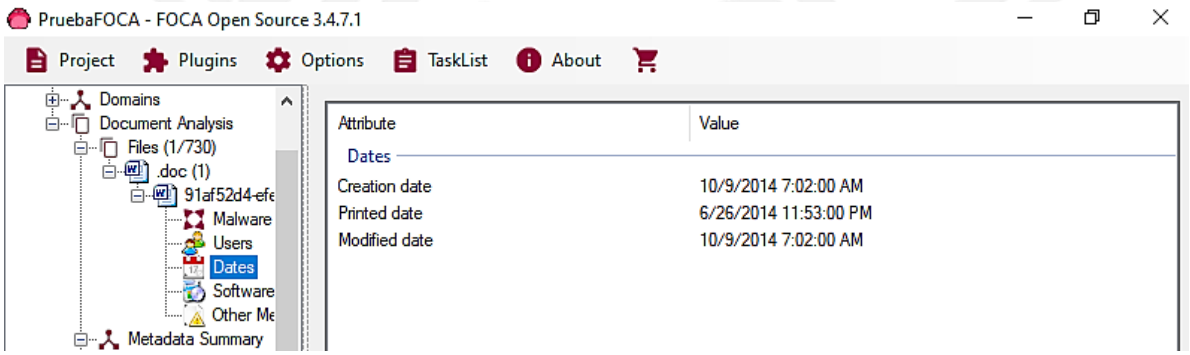


Figure 100. FOCA metadata by groups (II). Creation dates

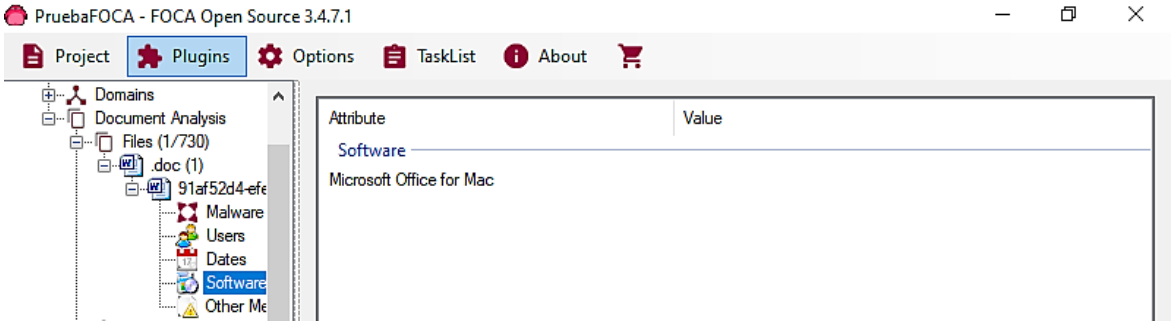


Figure 101. FOCA metadata by groups (III). Programs used to create the documents

If we now download a different document, we see how we have the same information, and **the tool collects and groups the entire web** in the "*Metadata Summary*" section depending on what it finds.

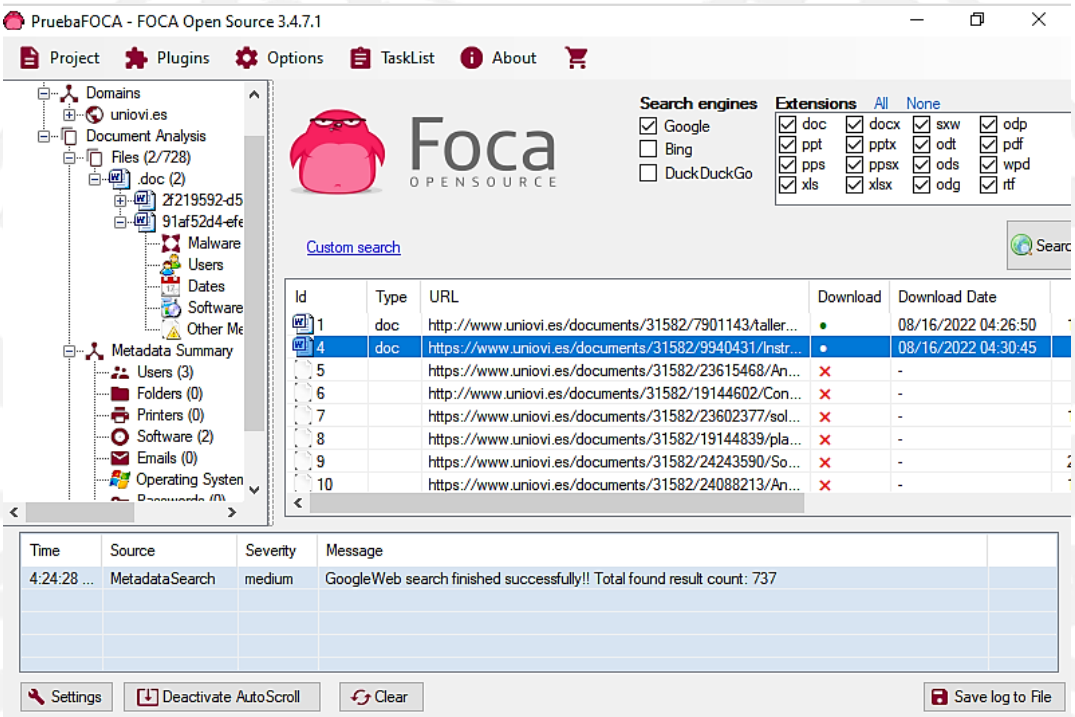


Figure 102. FOCA Metadata Collection

In the summary entries we can see the list of users, folders, emails, OS, etc. that are found thanks to the metadata of the analyzed web.

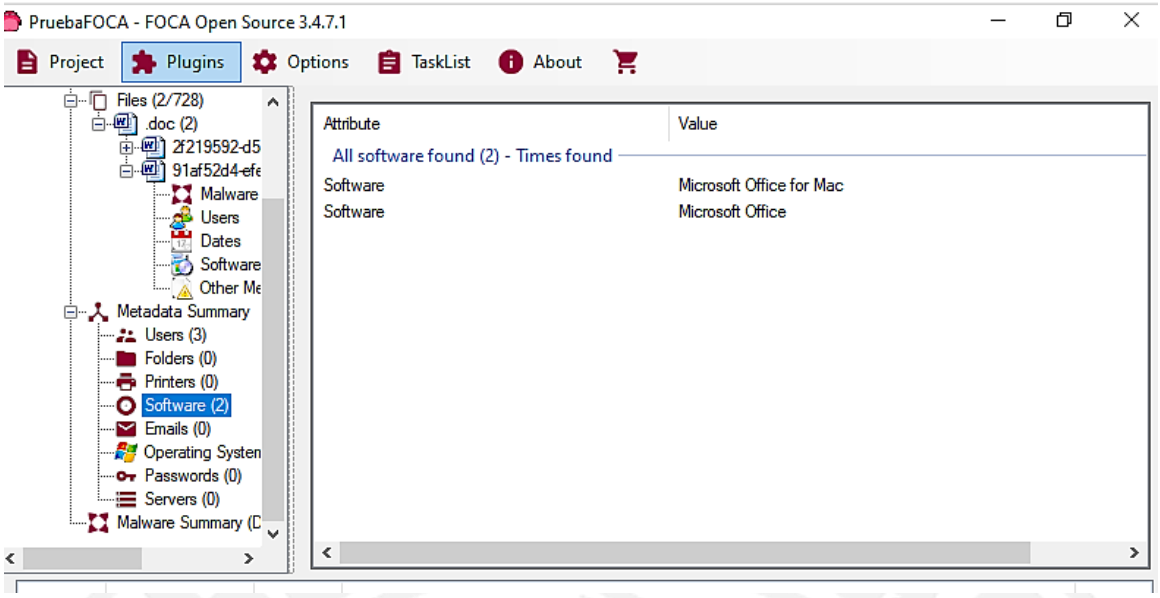


Figure 103. Aggregated data obtained by FOCA (I)

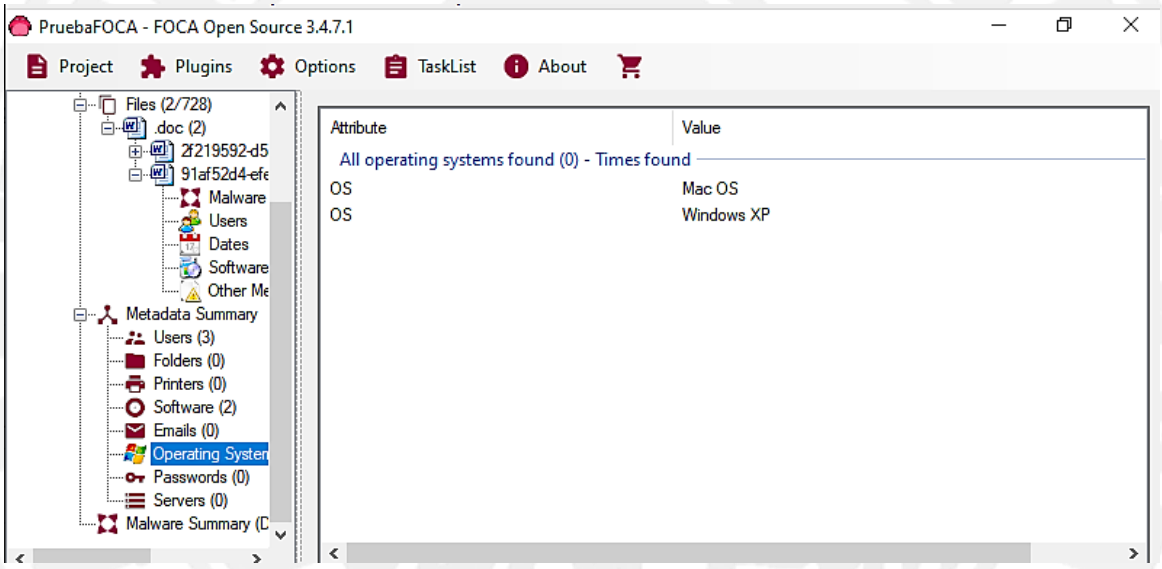


Figure 104. Aggregated data obtained by FOCA (II)

In the **options** we can finally enter the API keys of other external search engine services that we can use to improve our analysis, such as some of the search engines mentioned.

Data Mining for OSINT: Maltego

Maltego is a well-known **data mining tool** for OSINT, downloadable from https://www.paterva.com/web7/downloads.php. The objective of *Maltego* is **to represent the information you know about an entity** (person, machine, network ...) in a visual and easily understandable way, using a graph of relationships between different nodes. **Each node represents a discovered element of a certain type** (there are many types: domains, IPs, ...) and each type is represented by a descriptive image. The information in that graph can be constructed by automatic or manual procedures, or a combination of both.

With *Maltego* it is easy, for example, to present to anyone **a graph** that links one result with another, such as the RRSS profiles associated with the employees of a company (represented by their domain), their telephones, images, email addresses, etc. We can add the information as we discover it with our own investigations, or let the tool make a series of automated queries against various sources of information or data. Specifically, we can identify relationships between these entities:

- People
- Groups of people (social networks)
- Companies
- Organizations
- Websites
- Internet infrastructure
- Affiliations to different sites, events, groups, or organizations
- Documents and files
- More entities, as needed in each context (created by us based on text or incorporated from third parties since, as we will see, it is a modular tool)

Maltego has a paid version and a free ***Community Edition*** (CE) which is the only one we can use in this course because of the zero-expense policy we have. However, the CE version is very powerful and more than enough for this course and for any OSINT study that we do on a low/medium scale.

Installation

Installing *Maltego* is very simple, and you simply have to download it from the official website. As you can see in the following image, it is available for Windows, Linux, and Mac.

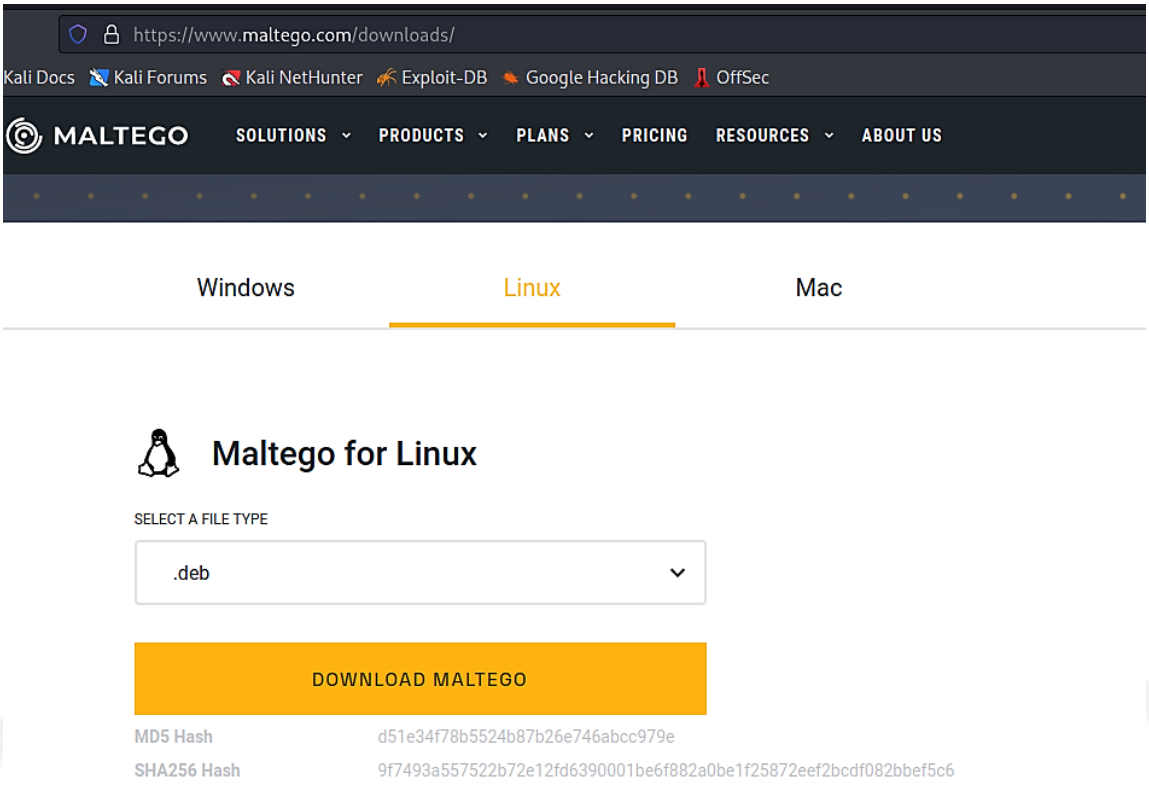


Figure 105. Download of *Maltego*

However, if you have a modern ***Kali Linux*** operating system with a graphical environment, it can be installed very easily **from its official package repository**, a method that is highly recommended because you will have the operating program very quickly and easily.

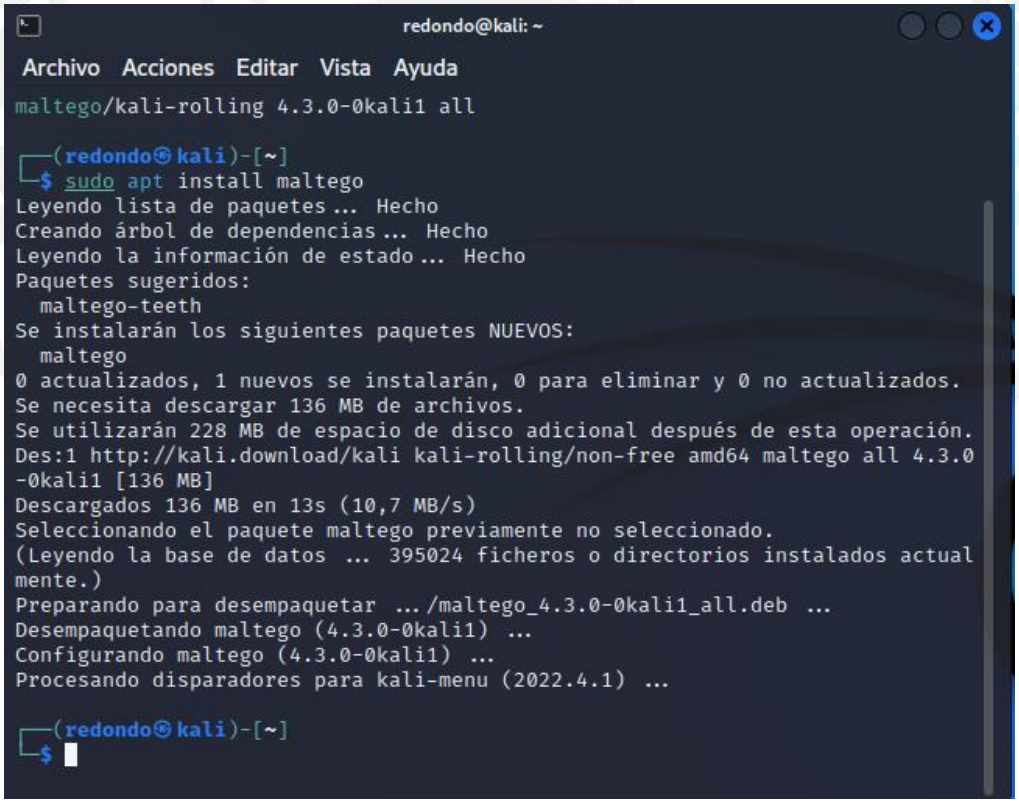


Figure 106. Installing *Maltego* from *Kali Linux* repositories

In order to use *Maltego* (including its free CE version, *Community Edition*) **it is necessary to create an account on the website of its manufacturer** (*Paterva*) through this link:

https://www.paterva.com/web7/community/community.php. The account is free and once done we can access the tool with the credentials we create. On the registration screen you can put that you are students and that your company is the University of Oviedo.

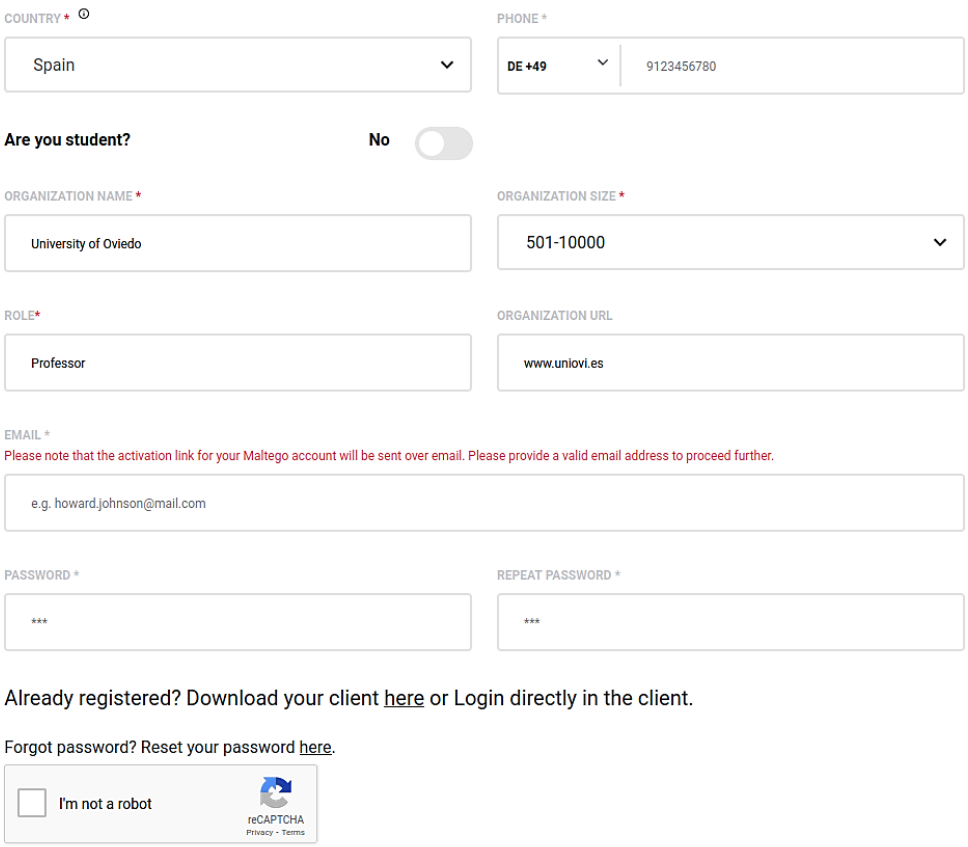


Figure 107. Registration on the *Paterva* website

The registration will send an activation link of the account to the email you provide, which could be a disposable one like those seen that *Paterva* does not prohibit, although in this case it is not necessary to be so cautious because what we are doing is something perfectly legal and normal. In any case you must make sure that it can be read, because it is necessary to click on it so that the account is considered correctly created.

First boot

Once the account is created we can start the tool and **select the CE version** (see image) when we are asked in the first boot which one we want to use. You could also use the ***CaseFile*** version, which is intended only to create graphics manually, without all the automated part that we are going to see. In this case it does not suit us, but for other uses it does (since it installs more types of base entities).

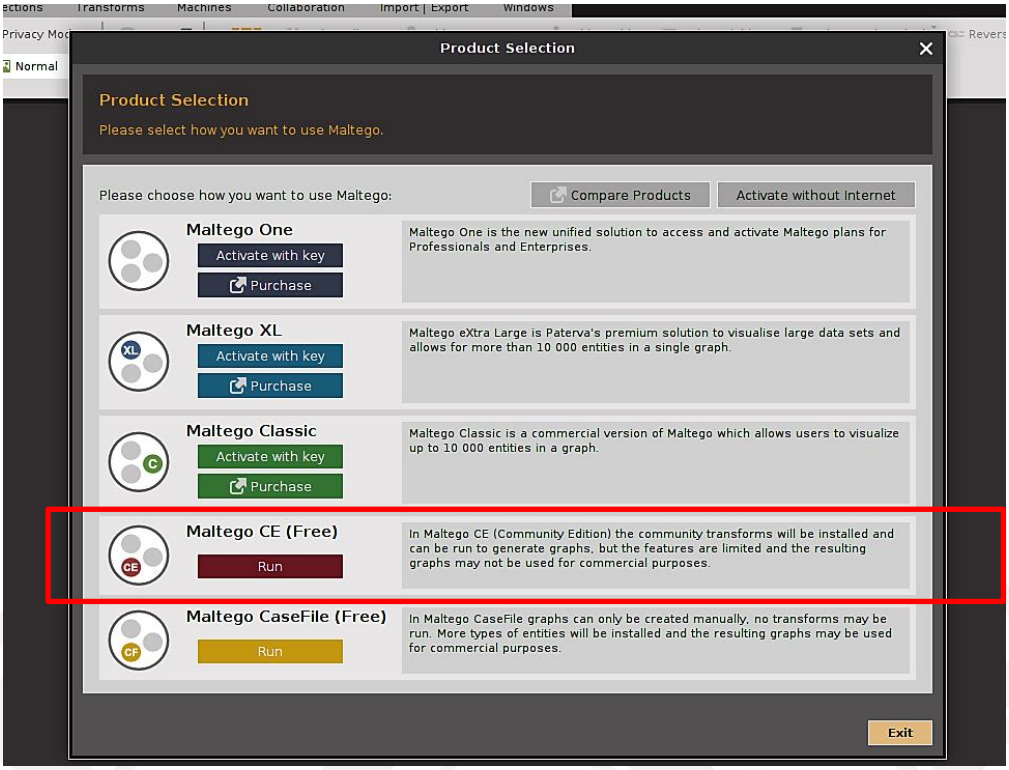


Figure 108. Free and paid versions of *Maltego*

Once the version is selected, it is time to enter the data of the *Paterva* account that we have just created, something that we should only have to do once.

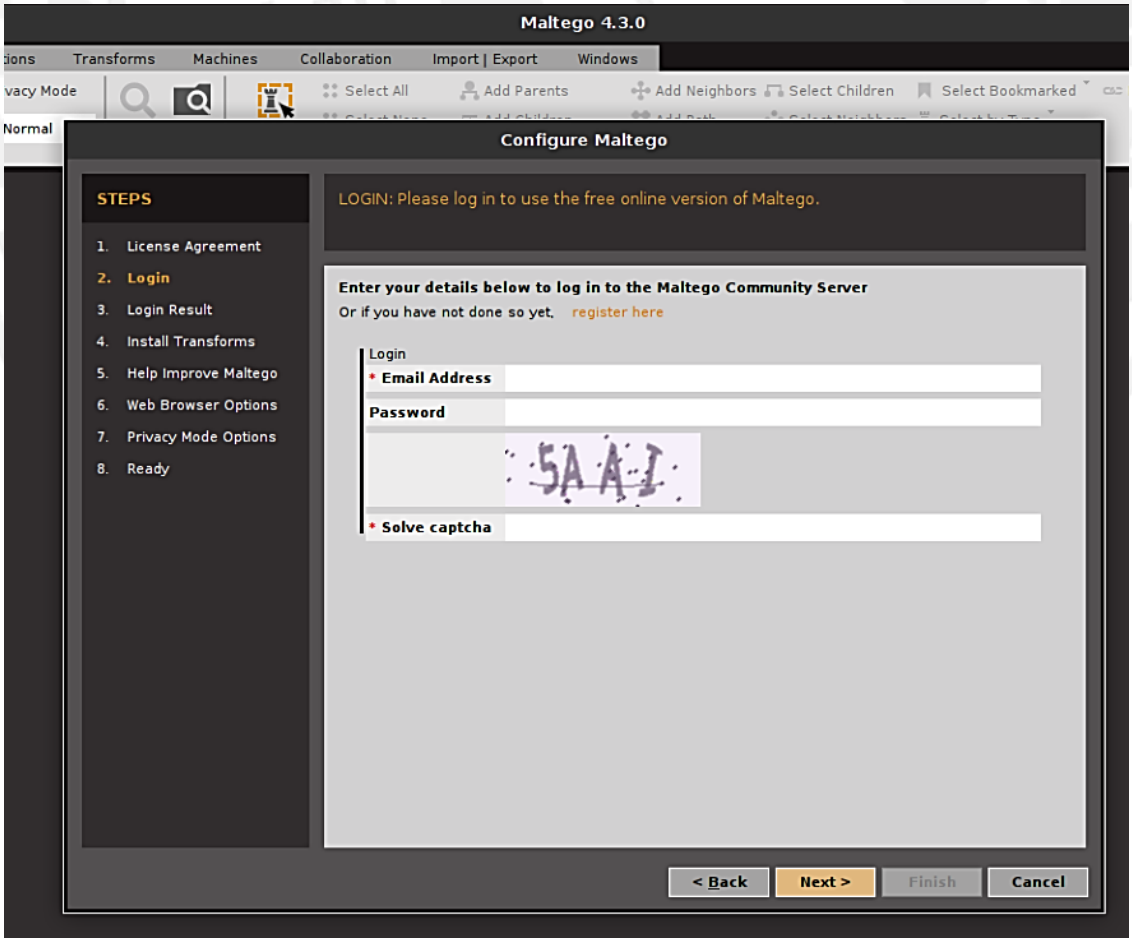


Figure 109. *Paterva* account data in *Maltego* on the first launch of the tool

If everything is correct *Maltego* will finish its installation by installing the necessary modules and we can move forward to finish its configuration.

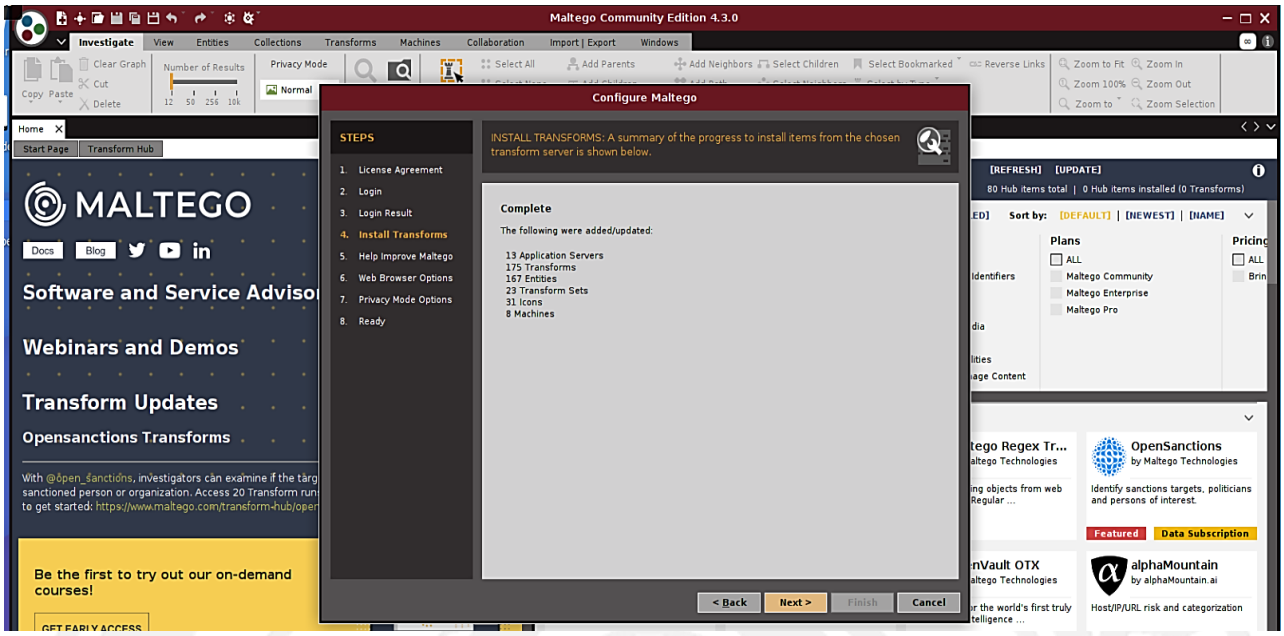


Figure 110. End of *Maltego* installation

Now we have a decision to make. *Maltego* can work in two modes: **Normal Privacy** and ***Stealth***.

- In the first mode our computer **will connect** with the investigated targets to download certain things, such as images and other data from their URLs. This generates traffic to the investigated from our PC, which can be analyzed to know who we are in case the target makes this type of investigation.
- The other mode **minimizes the traffic that our PC generates to the outside** (it does not generate traffic to the target, except in very specific cases), which is important in scenarios of serious real investigations (such as those that the police can do).

In the examples I am going to show **I have chosen the normal mode** because I am only going to make use of public information and I am not going to do anything wrong against any sensitive entity (come on, I am not a secret service or an attacker 😊). **The normal mode also ensures that as much data as possible is obtained**.

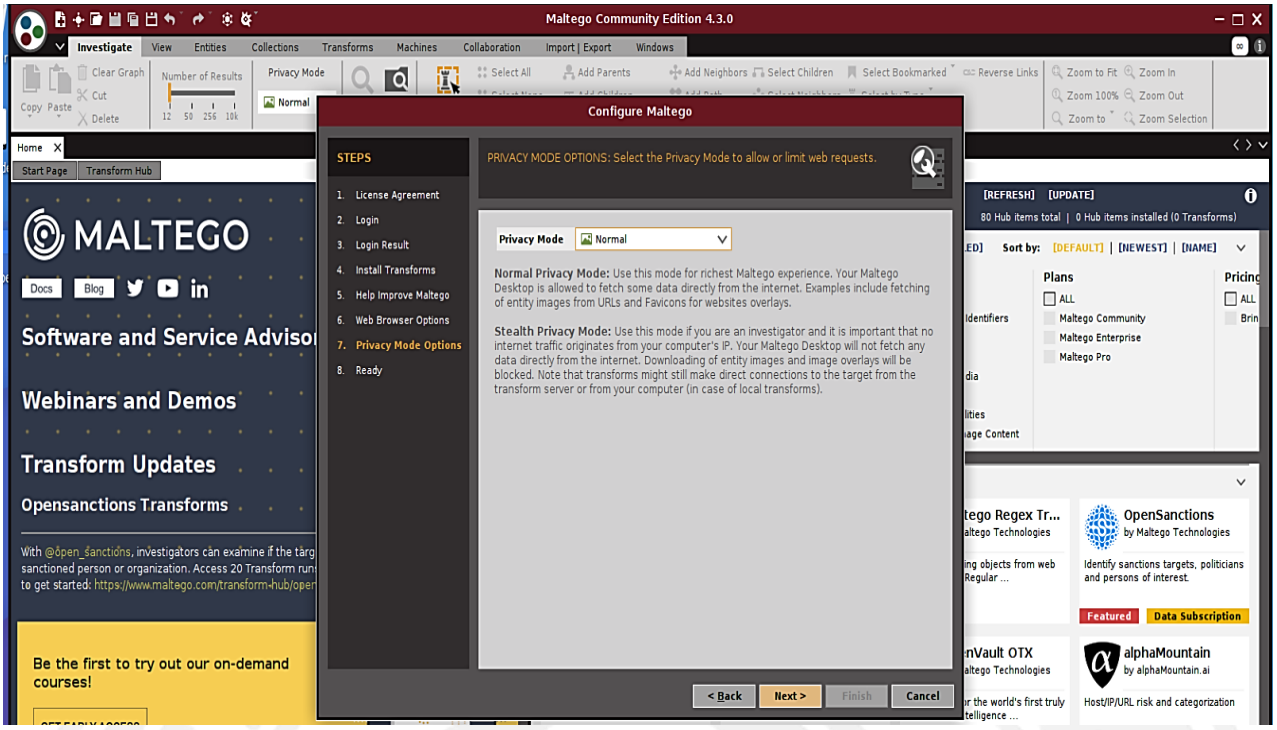


Figure 111. Operating modes of *Maltego*. Are you human or are you hacker? 😊

Once this initial configuration is done, we can use the tool to do our research. To start doing it we opted for the last option, and we were left with the welcome screen.

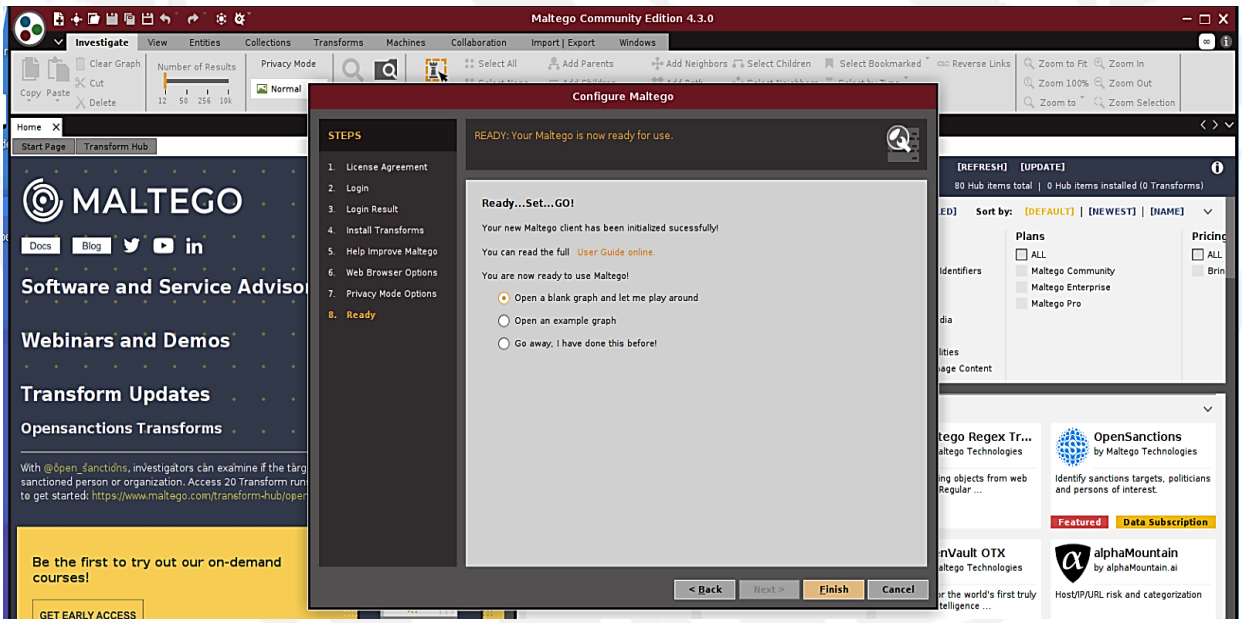


Figure 112. End of Maltego configuration

The initial screen shows the following content, which shows the different operations that we can do with the information available in the community version. These operations may change with the version we install.

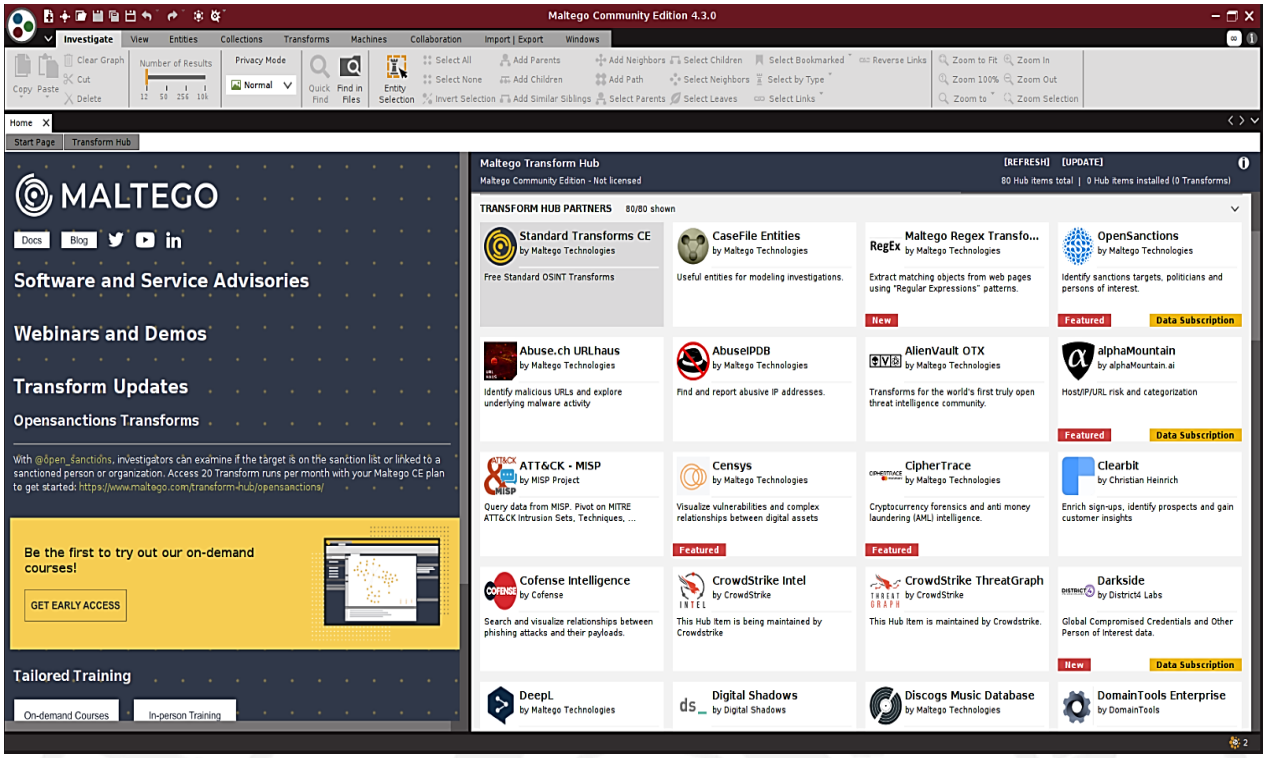


Figure 113. *Maltego* welcome screen

Creating a project

***Maltego* works with projects**. Each project keeps the relationships that we can establish on an entity, and in this way we can maintain multiple projects for several entities under study. For now, we will create a new project (called "*Graph*" in the tool) as follows:

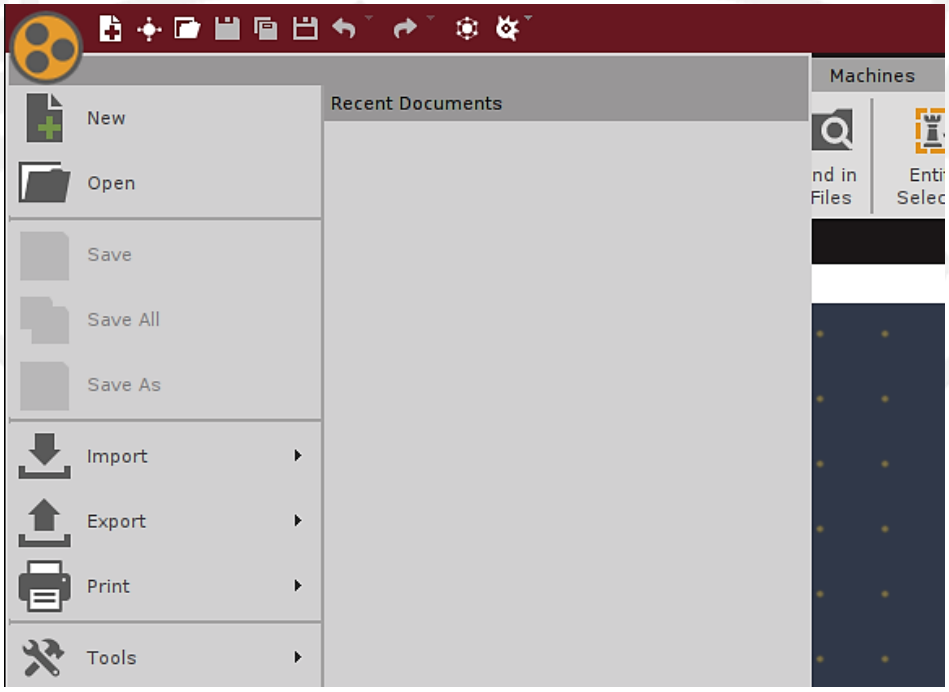


Figure 114. Creating a project in *Maltego*

Once created, it will open a blank window and in the left bar we find a window called ***Entity Palette***.

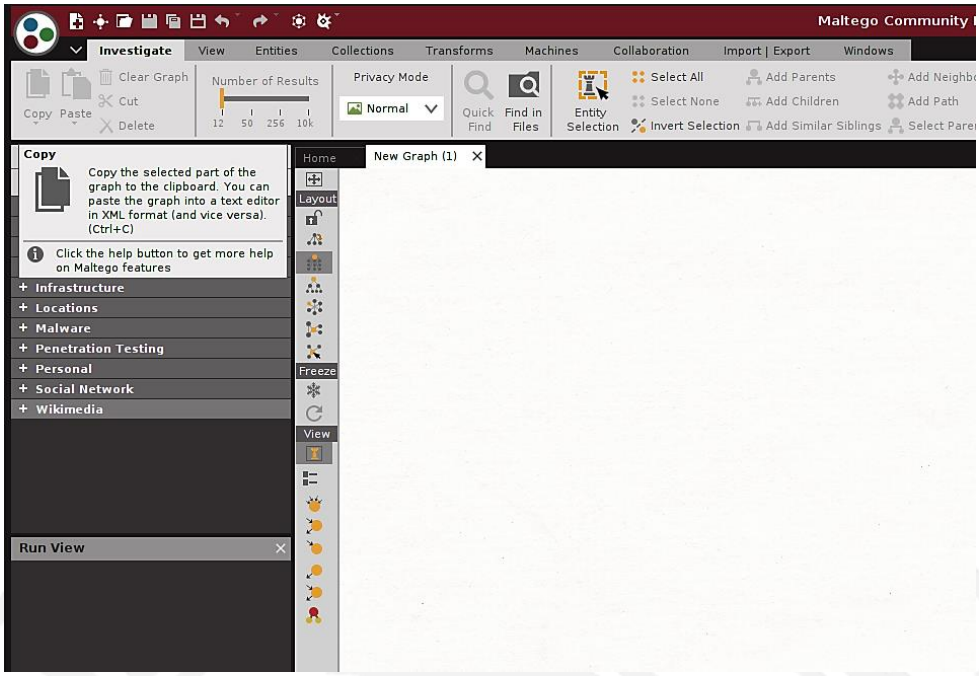


Figure 115. *Maltego* entity palette

To start working, all we have to do is drag an entity and drop it into the blank ***Graph***. In this case we choose an entity of type "*Domain*", which will represent any domain on the Internet.

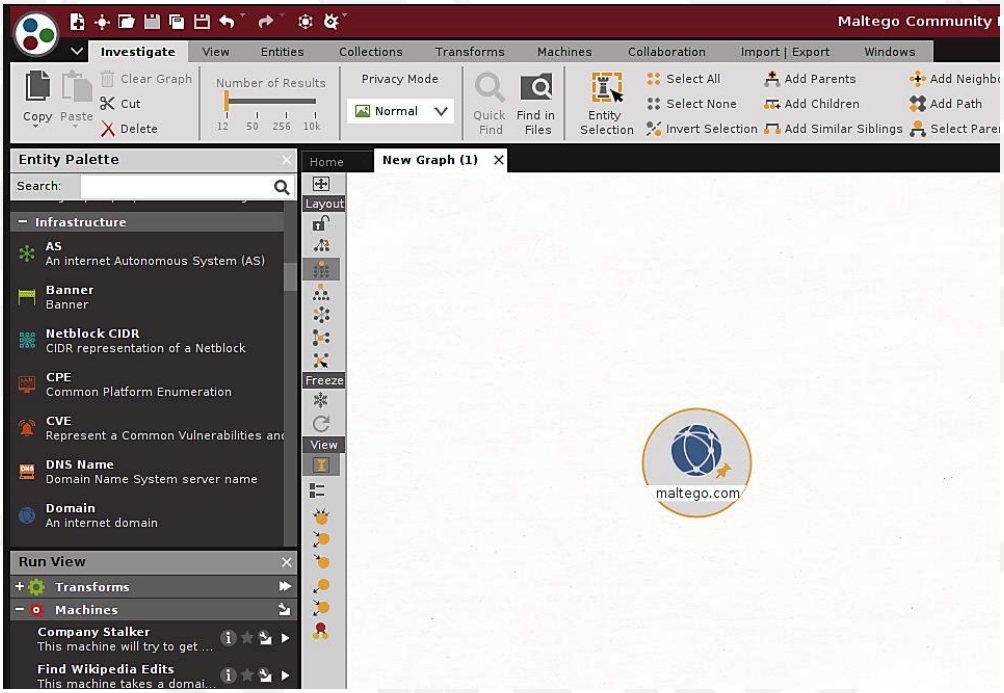


Figure 116. *Maltego* graph with an added entity

To **edit** any entity, we can double click on it or go to the right panel in the "*Property view*" section and edit the content that appears as an example, in this case `maltego.com`. It is very similar to a visual editor for creating application GUIs. Now we can change that domain for whatever we want, and the game begins! 😊

Figure 117. Property view of *Maltego*

Automated information search

As an example, we are going to change the previous domain to `uniovi.es` and we are going to tell *Maltego* to automatically do a **search for public information about it**. To do that we give the right button on the entity, and we see the search options for information of several types that come out. Each of these options is called **transformations** (***Transforms***), which represent the different searches supported by *Maltego* at the moment for the selected entity type. **Obviously, each entity has a different set of options**. For a domain we have these:

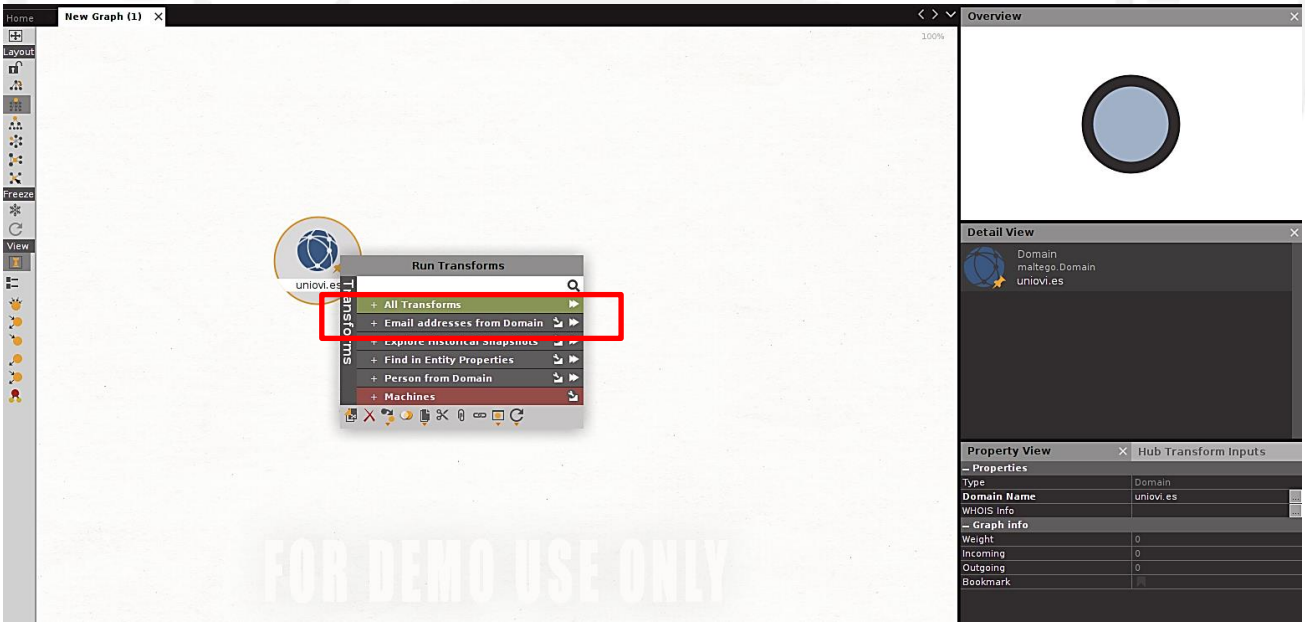


Figure 118. Transforms of a *Maltego* "*Domain*" entity

If we click on the name of each **transform** we can see in detail each of the operations that it does. This allows us to apply them one by one (by clicking on the icon of the arrow next to the name) and make a detailed and ordered search of those aspects that interest us most in each entity, in order to have a graph only of those things that interest us (from the list of things we saw in the "Manual OSINT" section). However, we can also go "kamikaze" and click on the arrow of the category "***All transforms***", putting the tool to work

against the ***Domain*** (or the entity that we have put) with all the transformations that know about it, to see what happens. A friend of mine calls this "*On my command, anger and Maltego*". 😊

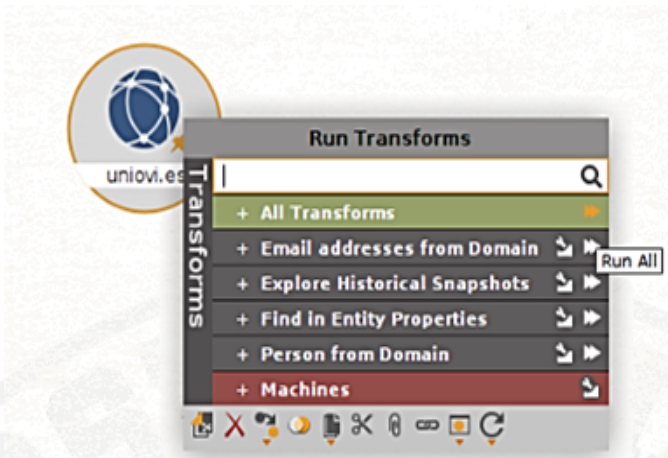


Figure 119. Starting all the transformations of an entity in *Maltego*

Therefore, **searches in Maltego are called transformations** and the number of results we will obtain will depend on the version of the tool we use. Our CE version has the ability to perform an analysis on 10,000 entities on a chart and can return up to 12 entities per transformation. As we said, to make a first test we use "***All transforms***" to the `uniovi.es` domain, getting a huge number of results that we can go on to analyze. Of course, some of these transformations need you to specify or confirm values of parameters that control their operation, which will ask us before doing the search. In this case, for example, start and end dates are being asked for the search for the domain on the ***Wayback Machine***, which we can leave with its default values.

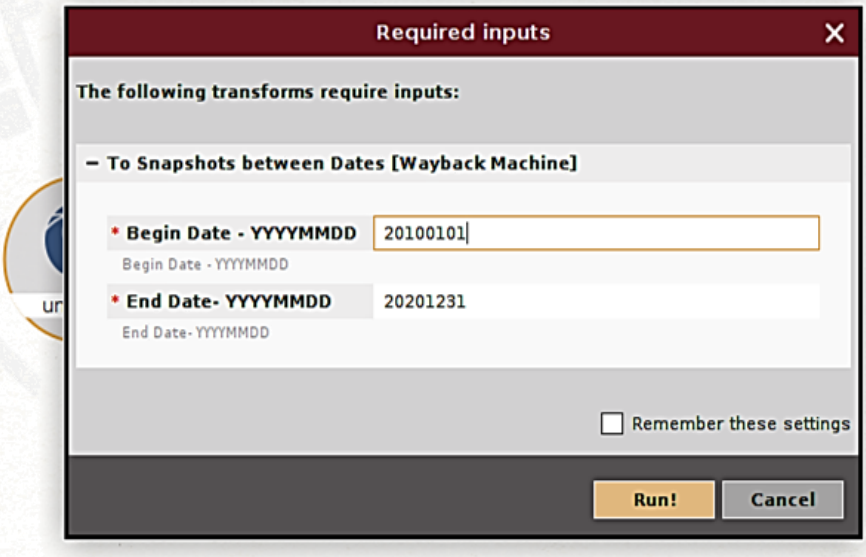


Figure 120. Request parameters of a *Transform*

Sometimes we can find that **one of the Transforms fails** for some reason at runtime (because of bugs they may have), but normally they are independent entities from each other and the fact that one fails does not imply that the others do: we will have results regardless of it.

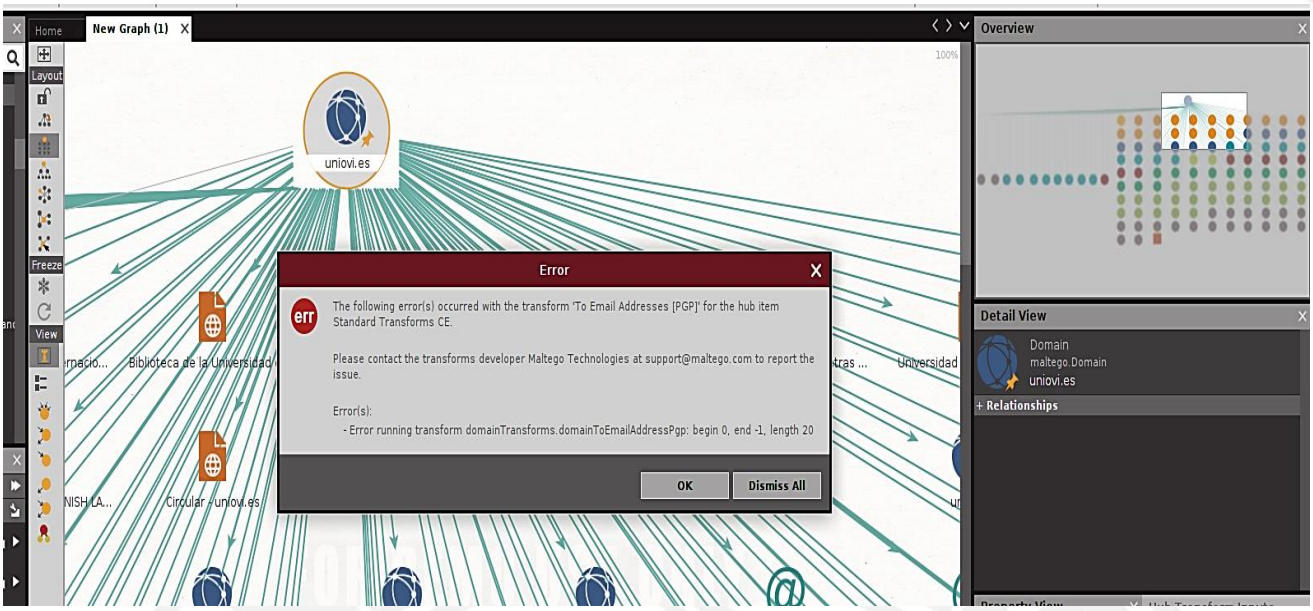


Figure 121. *Maltego Transform* runtime failure

At the end of the search, we see that **a huge number of entities** of other types appear associated with the domain `uniovi.es`: documents, machines, IPs, e-mails, people ... and all of them can be investigated individually using the same technique of **right click + *Transform*** one by one, with which you can verify that the possibilities of "exploration" of this tool are enormous.

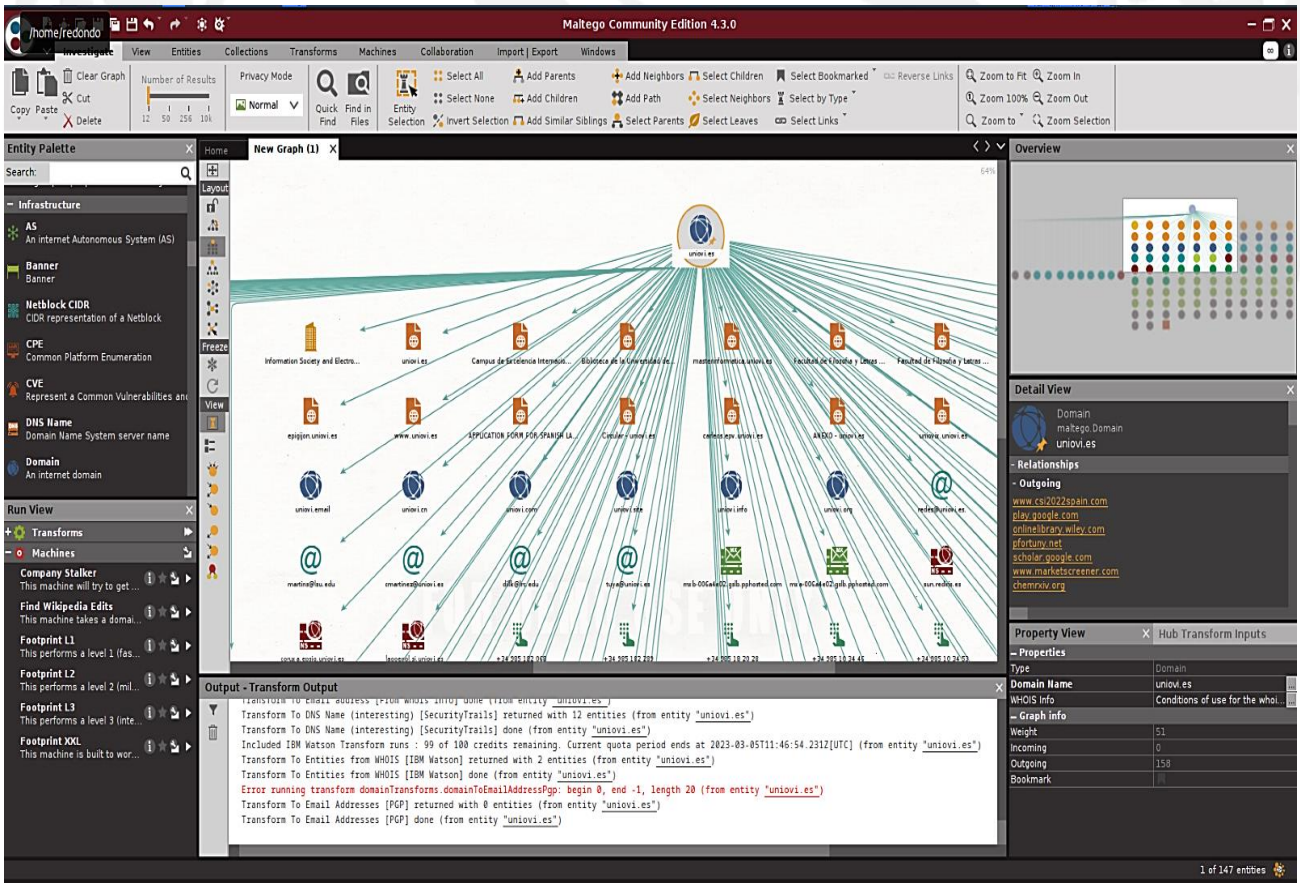


Figure 122. Display domain information thanks to automated *Transforms*

These searches are possible **thanks to the different techniques that *Maltego* uses to collect information**, which frees us from doing them manually as we saw in the previous section. These techniques are similar to those mentioned elsewhere in the document: `whois` a domain, ask DNS servers, use various search

engines, search different social networks such as *Facebook*, *Twitter*, *Instagram*, etc., but the tool does it for us. It is a way to obtain "massive" information when we do not have to "do precise bombing" and we need to have as much information as possible to start studying the objective.

The graph, which can be huge, is navigable and we must go through it little by little in search of interesting things, such as telephones and other websites associated with the investigated domain, as we see in the following image.

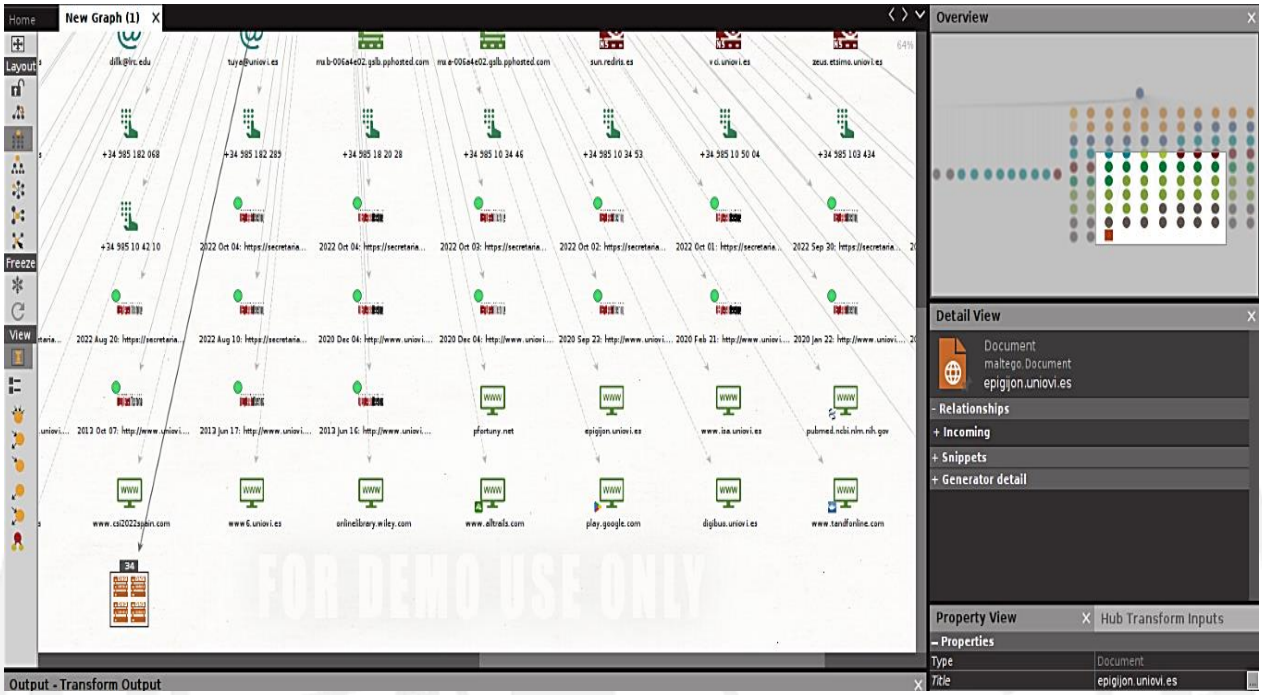


Figure 123. Miscellaneous entities discovered by *Maltego's Transforms*

Search among what you found 😊

As we said before, once a set of information is obtained, we can continue to make transformations to any of the entities that represent data discovered by *Maltego*. We therefore see how **the discovery process is completely recursive**. Before we did it on a domain name, which led us to discover a subdomain associated with that name. Let us use this entity and analyze it again with "***All Transforms***". I insist that this is a "brute" approach to the problem, normally one goes “finer” and more detailed, but here we have come to pla... experiment 😊:

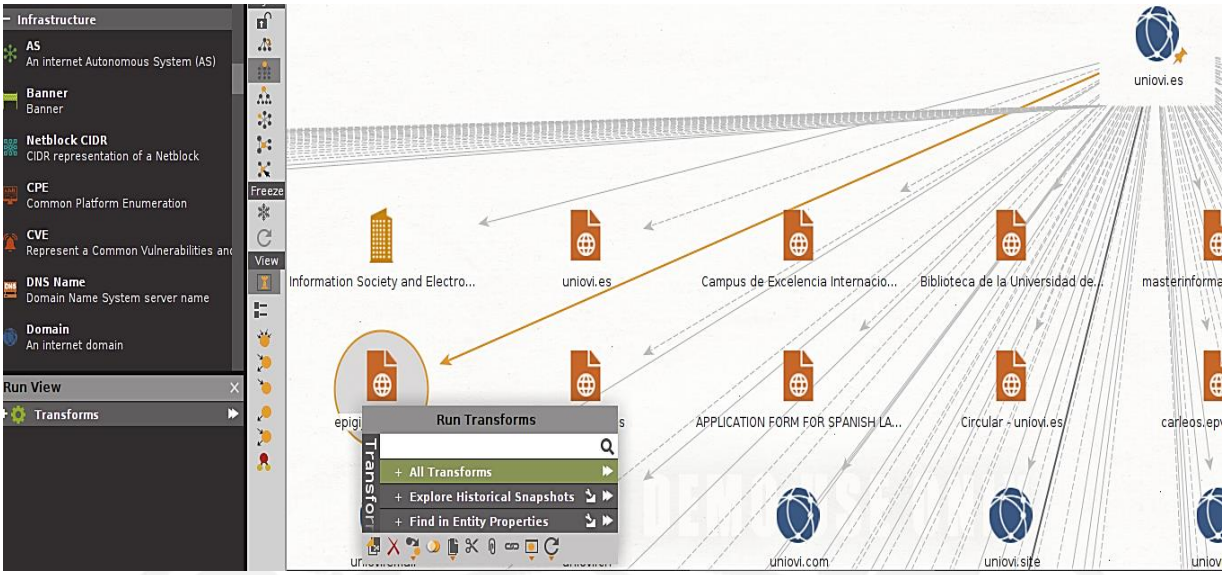


Figure 124. Running all Transforms on one entity from *Maltego's* first search

This gives us a new set of results:

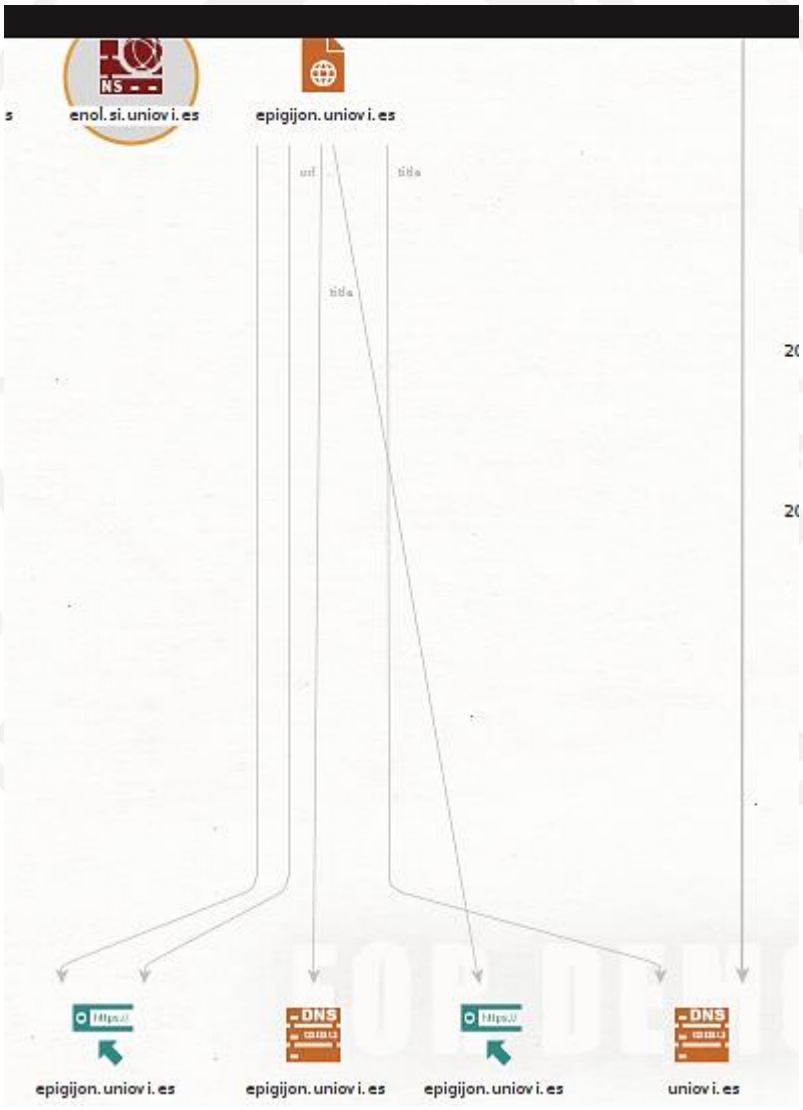


Figure 125. *Maltego* results after running the *Transforms* of a subdomain

Now we get a ***Website*** entity associated with this domain, on which we proceed to do the same. Again, we are asked for parameters for the *Transforms* that need it, and now they appear more than before since the

Transforms that are executed now are different from those of the first time (***Website*** entity vs ***Domain*** entity). In this test we leave the default values again.

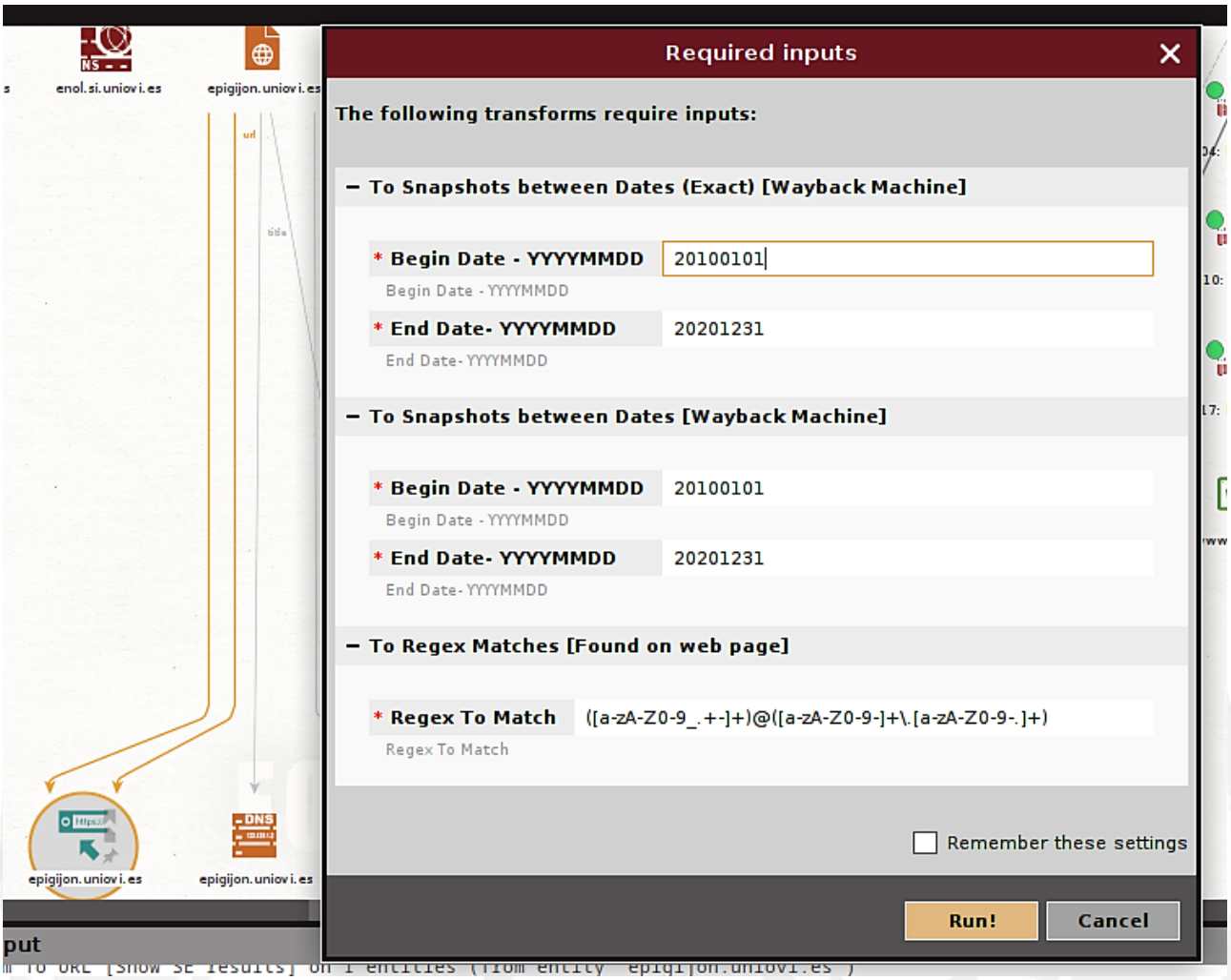


Figure 126. Parameters of the *Transforms* of a *Website* entity

We see that now *Maltego* obtains other entities, such as **the framework used to create the web and some others of dubious reliability** such as the use of payments in pounds or yen. As we will say later, **not everything that *Maltego* obtains is reliable**, you need to analyze each entity manually (with the techniques of the previous section) to see whether or not this information is reliable and if it really gives us something. Of course, **we can delete everything that is useless to make our graph more manageable**.

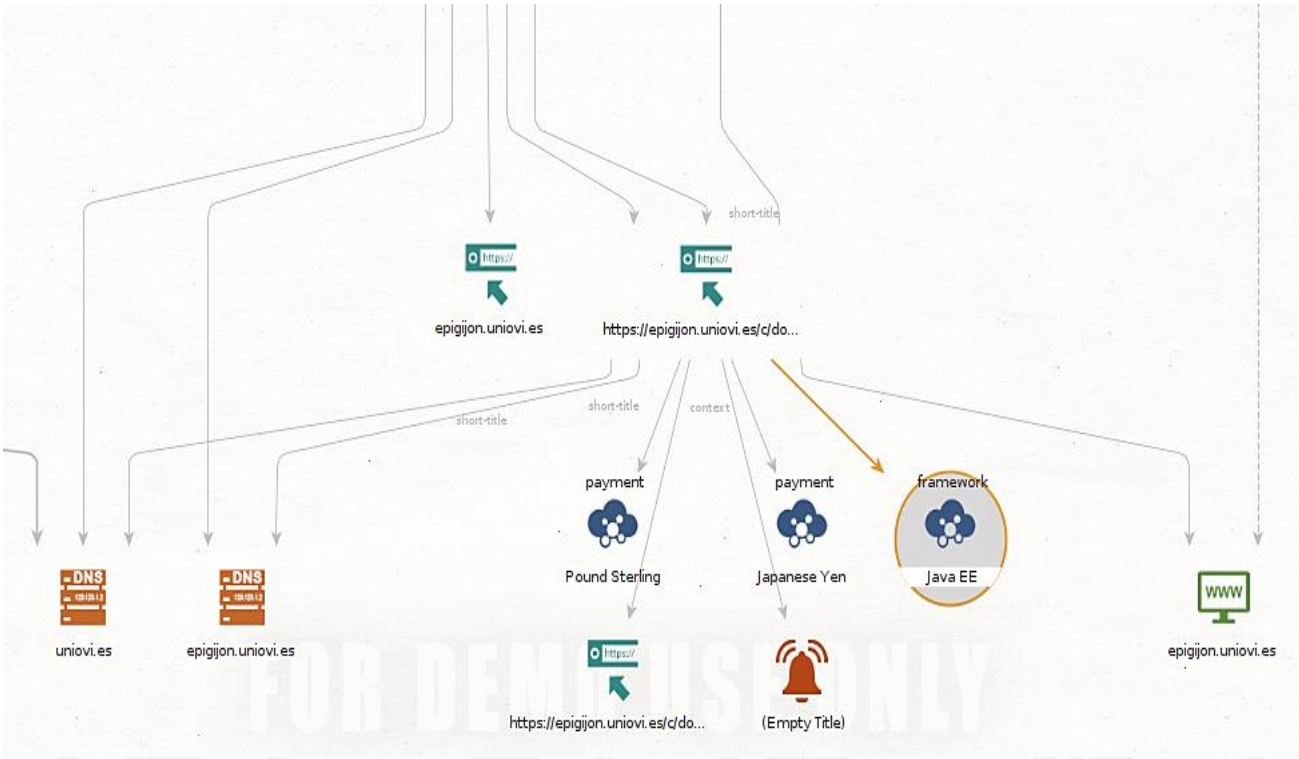


Figure 127. Website analysis in *Maltego*

If we continue analyzing we can reach the DNS server that is responsible for serving the IP of the scanned web and an IP of this, which can then be explored by active techniques such as those we see in the course (e.g.: `nmap`). As we see, we can get to know many things (high and low level) of any entity explored if we insist enough 😊.

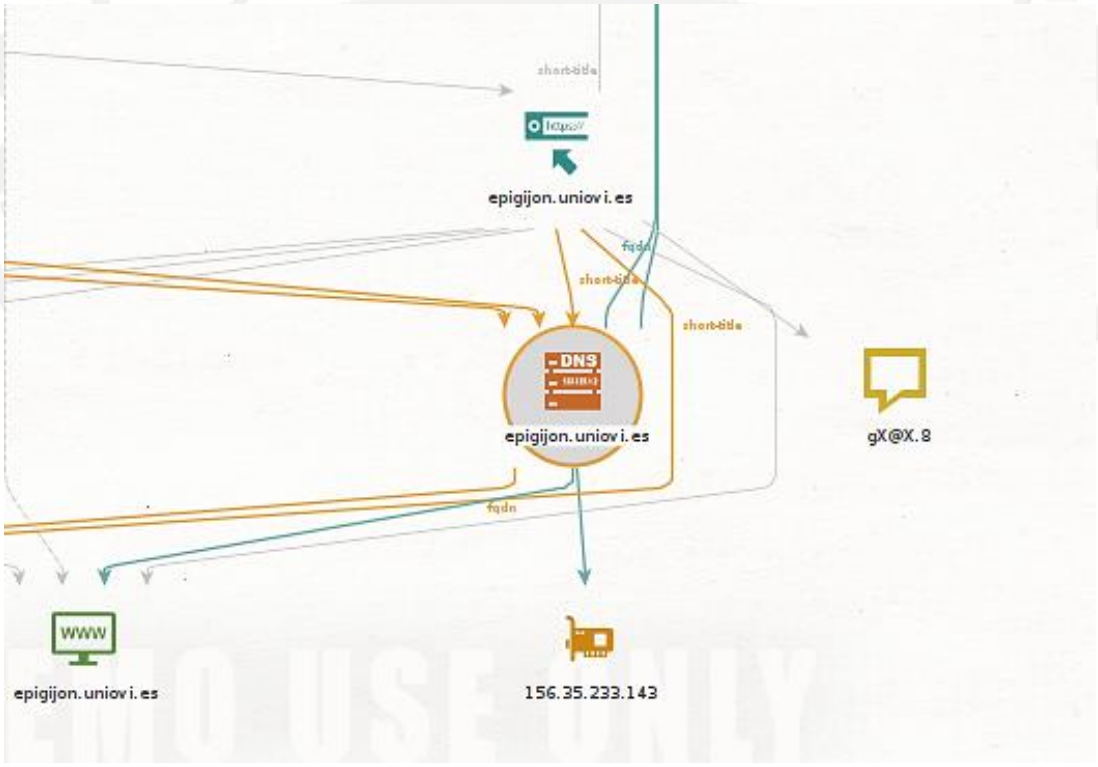


Figure 128. DNS and IP of a website obtained with *Maltego*

We see therefore how this process can be extended as much as necessary, being a way to coordinate all the information discovered of an entity through "brute" transformations (as we did), more specific (if we know what we are looking for) or providing data that **we can discover through our "manual" investigations**

with other tools given in this document in the previous section or with third-party transformations that we will see later.

Before continuing, it is important to highlight one thing. *Maltego* is a very powerful automatic tool for data mining, but we saw that, like any automatic tool, it can return information **that is not entirely reliable or directly false**. That is why we have to know how to get information with other more "manual" tools: not only is it possible that we do not have *Maltego* at hand, but we also have to confirm that the information it returns to us is true in case it is useful for something. It therefore makes perfect sense that if *Maltego* discovers an IP for example, then we go to sites like ***IPInfo*** (https://ipinfo.io/) to confirm that the data is 100% accurate. Do not trust everything *Maltego* tells you by default (the probability that it is correct is high, but not 100%).

Save your searches

As, obviously, all this search information can be too much to "digest" in a single work session, we can save the graph in a file to examine it several times in this way:

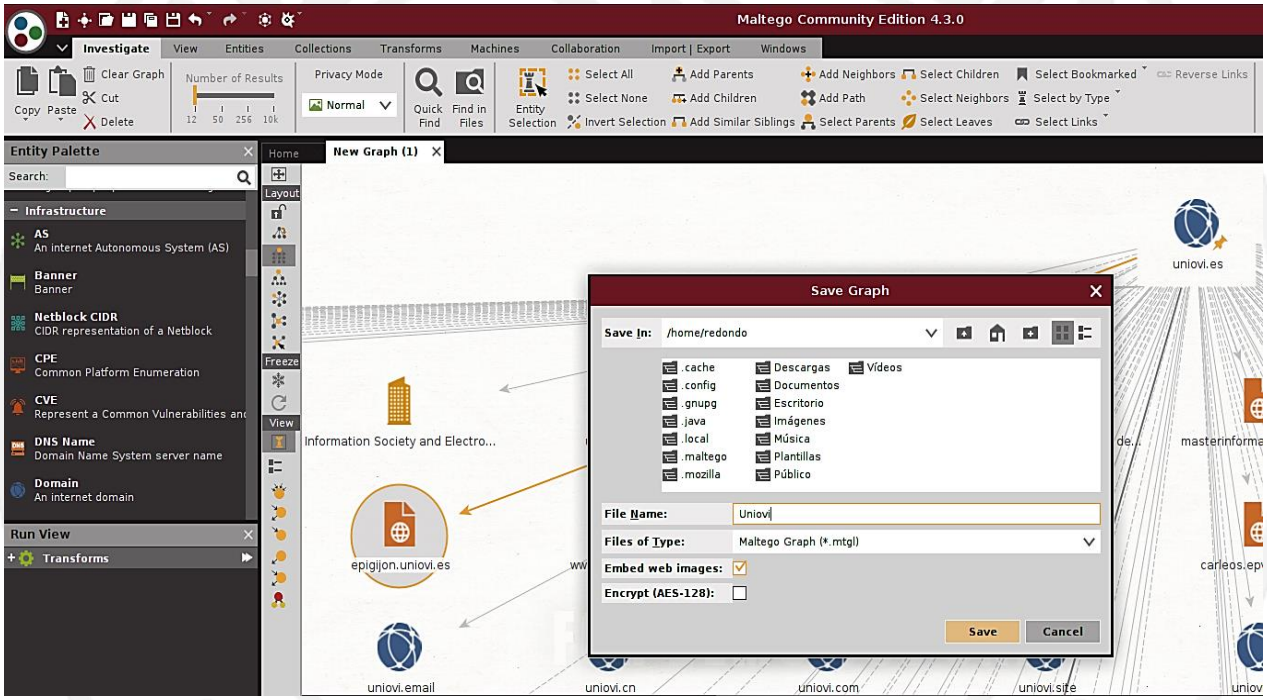


Figure 129. Save a graph with *Maltego* searches

Detail of a *Maltego* entity

Once this general exploration is done, the next objective is **to review each of these results and their entities**, choose the ones that interest us most for what we need to do, **and double click on them to see their details**. The following screen is a typical tab of one of those entities:

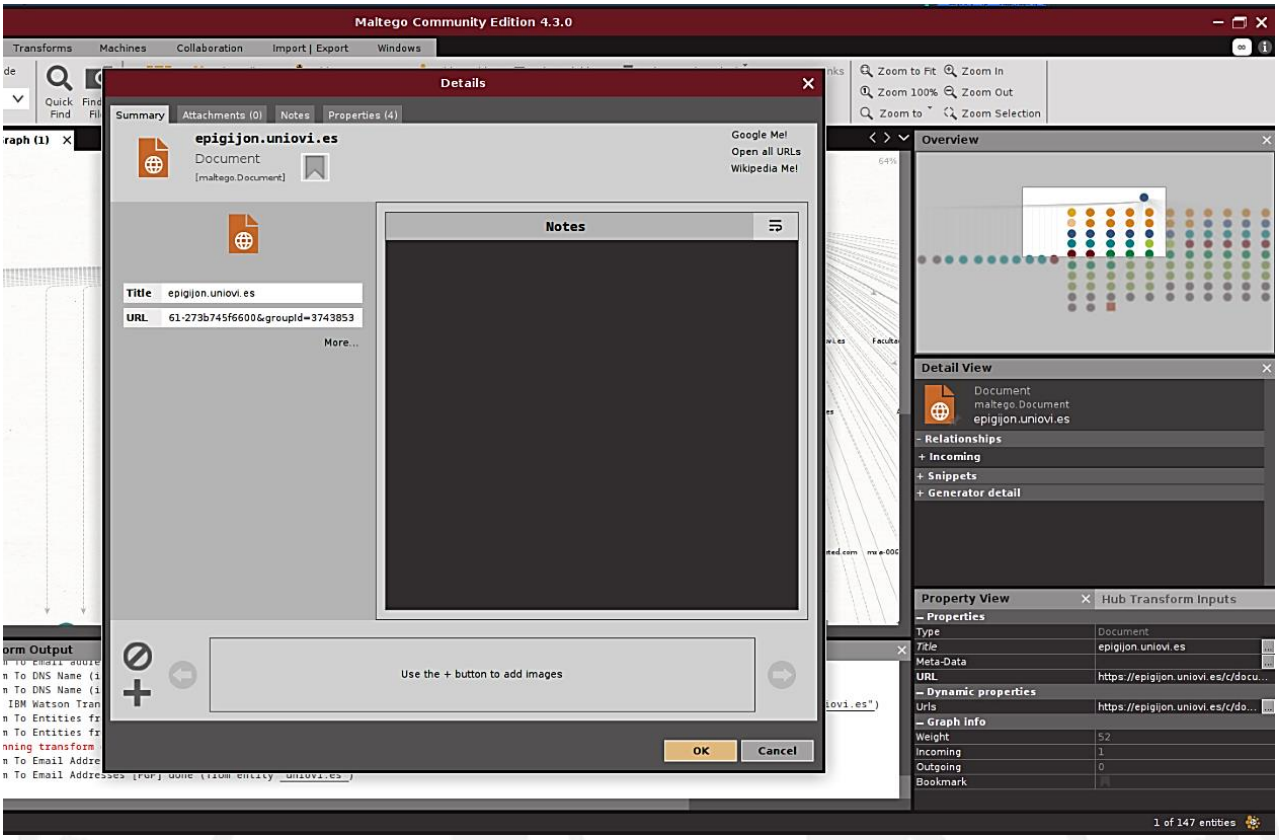


Figure 130. Details of a Maltego entity

The details of an entity are divided into:

- **First tab**
  - A **summary** (at the top)
  - The option to do a ***Google* search**, open in the browser (although only because the analyzed entity is a domain) or search for information about it in ***Wikipedia***, if we do not know exactly what the entity we have found is.
  - Option to **add images** associated with the entity to use later in our research (e.g., screenshot of a website, profile picture...). Keep in mind that this graph **in real situations will have to be seen by many people** and the more information we can put about each entity, the easier it will be for people to understand it or check what it is with the naked eye. The more you think about others who are going to use this "map" you are building, the better. Usability is also important here! 😊
  - A **notes section**, to write what we see fit as we investigate the entity, especially something important that we want to remember later (to ourselves or to others).
- **The second tab** is the attachment tab. This is used if we want to attach something, such as reports of the entity or person, files that have been generated elsewhere, captures of *Tweets* or messages in general RRSS, for example. In other words, **proof and evidence** about the entity that will then be useful for something.
- **The third tab** is for adding extended notes about the entity. Sometimes you find things that require a detailed description or leave notes/questions for other team members who are examining the chart entities. **Remember**: this is a research tool.
- **The last tab** is the properties tab, where there is often information about ports, images, URLs, metadata, or other background related to the search that *Maltego* has found.

As you can see, a thorough professional investigation with *Maltego* requires patience, technical knowledge (although not especially advanced), attention to detail, properly documenting the findings by combining automatically obtained elements with others obtained manually and, surely, the collaboration of several people with different profiles. But it is, without a doubt, exciting 😊. Cibermaruja++.

Creating graphs manually

Previously we have said that with *Maltego* we can represent the information that we find with the "manual" analysis that we do with OSINT procedures. For this, the tool supports the creation of graphs manually. *Why do we want to do that?* Well, because the information can also be **found or inferred** manually (as we saw in the previous section), we can make assumptions (which is good to document in the detail of the entity), or through other tools.

Remember that in **Seminar 1** we saw many tools that are used to obtain more or less automatically information from people, machines and code that would join the one we have and that *Maltego* probably cannot find. All this data we have to add to the graph we have, dragging, dropping, and linking the found entities "manually" with the "automatic" ones and thus **enrich the final result**.

To do this we must drag the entities we need, edit them with the information we have discovered, direct the cursor to the object we want to link and move to another generating a "link" establishing some type of identification of that relationship, as seen in this example:

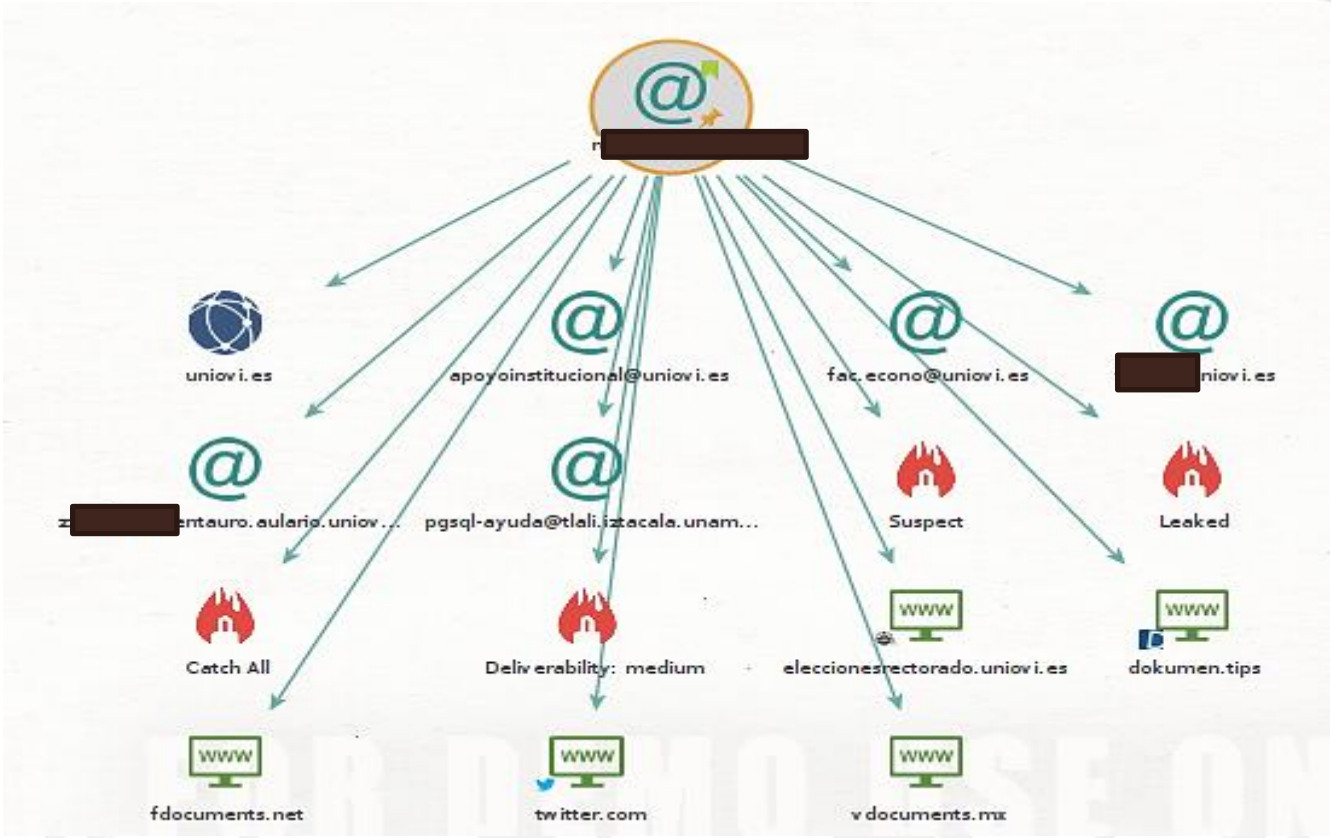


Figure 131. A person's set of information combining automatic transformations with manually obtained information

By doing this we can visualize the information obtained in an orderly manner and complete our research with what we collect (and verify). In addition, everything added manually can be combined with the automated part of the tool: if we select some of the added entities (**several can be selected by pressing capital letters**), we can apply "*All transforms*" again on all the selected ones and more data can come out. It is possible that in these cases the tool asks us again for parameters to apply to the transformations before executing them. For example, if you are going to use a search engine they can ask us for a specific one that we want to use (Ex: www.google.com). In these cases, we can get interesting new information, **such as if an email address is filtered** in a known stolen email database, as seen in the following image.

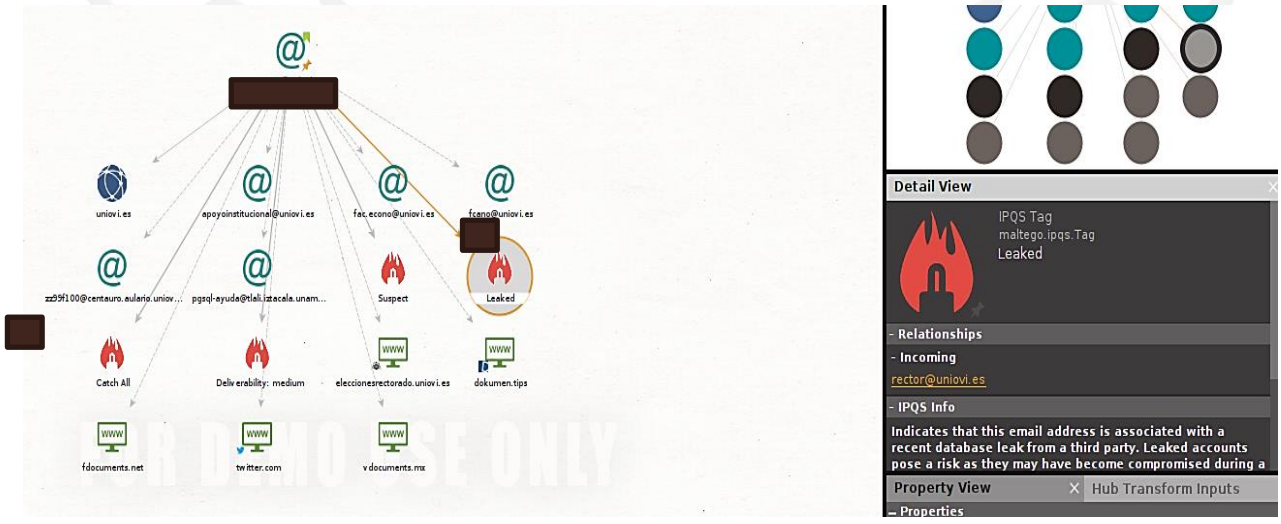


Figure 132. Email address leaked in a database detected by Maltego (do not worry, it is not a dangerous leak, the key has been changed since then 😊)

We see how the tool is able to detect important things, such as if the address is suspicious of sending spam, if it handles errors inappropriately... as I said, a *Maltego* graph can tell a very interesting story that is worth reading 😉.

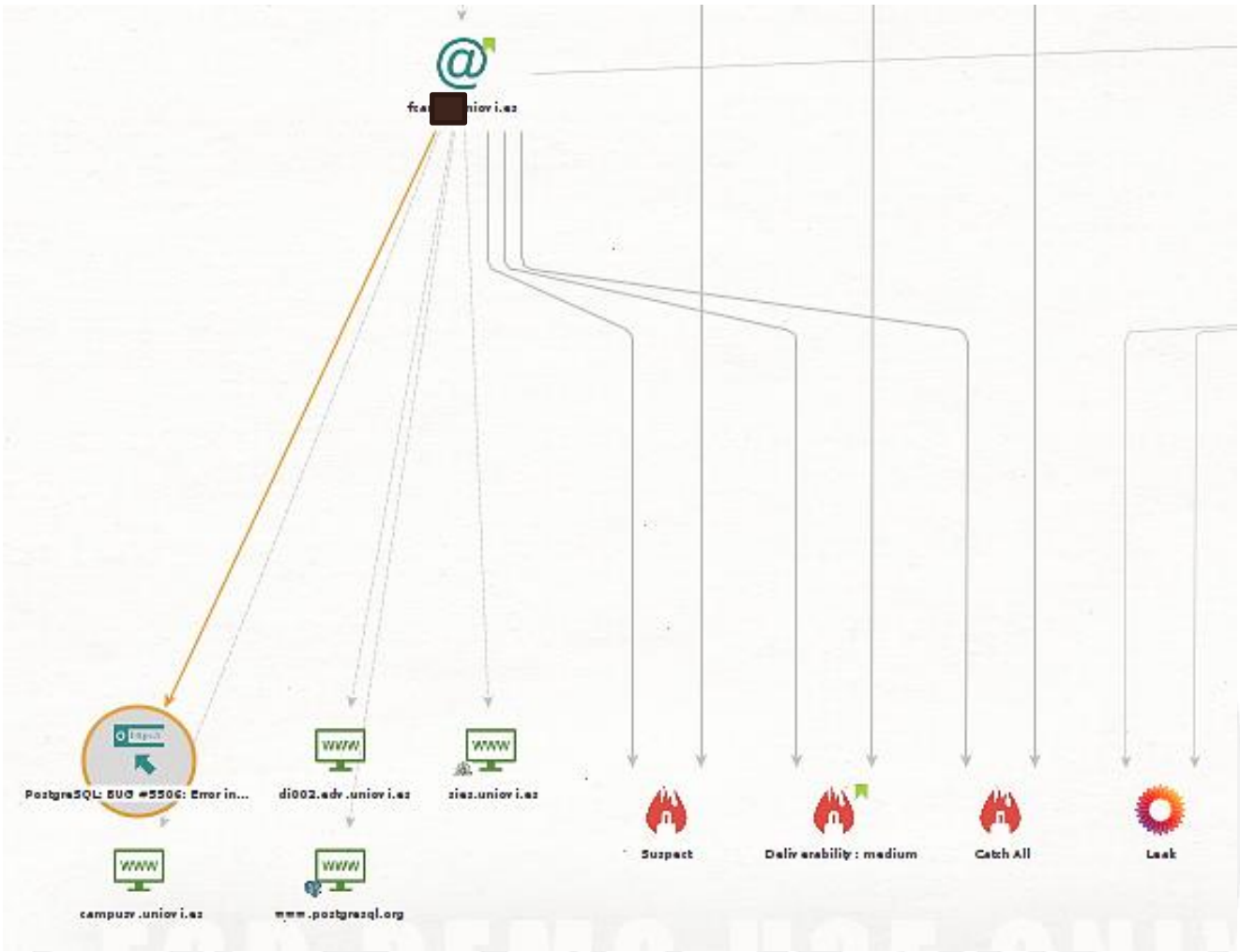


Figure 133. *Maltego* discovered interesting things about an email address

In this example we manually add a ***Person*** entity, put a known name, and do an automatic analysis to get more emails, location, DNS, and aliases, among others. We also have social media profiles, which we will talk about in the next section.

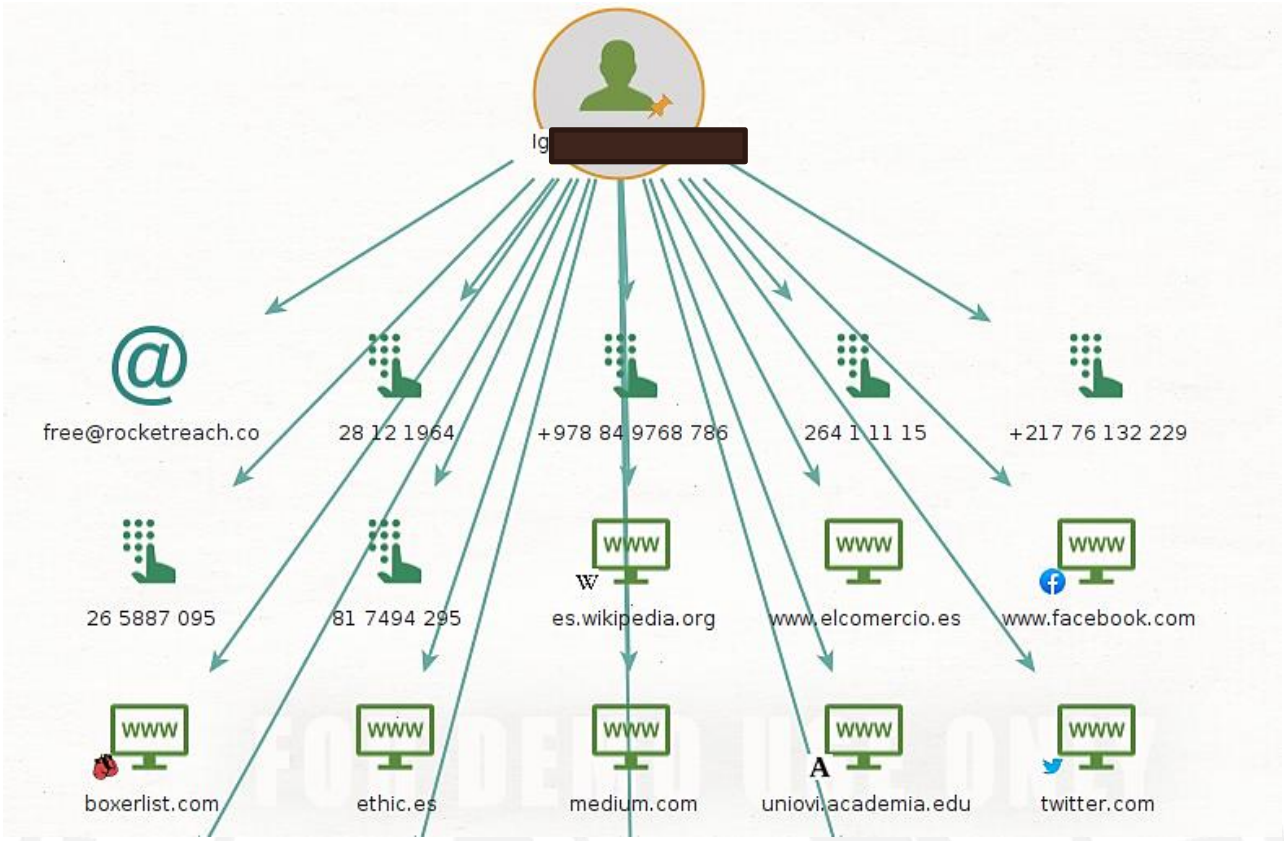


Figure 134. Typical analysis of a person's name made by *Maltego*

Of course, doing a little ***egosurfing*** (looking for oneself) is good to reveal if we are giving out too much public information. This is my case, where only appears a) false information (or other people who have my name) or information that I have public for my work. So, all good 😊.

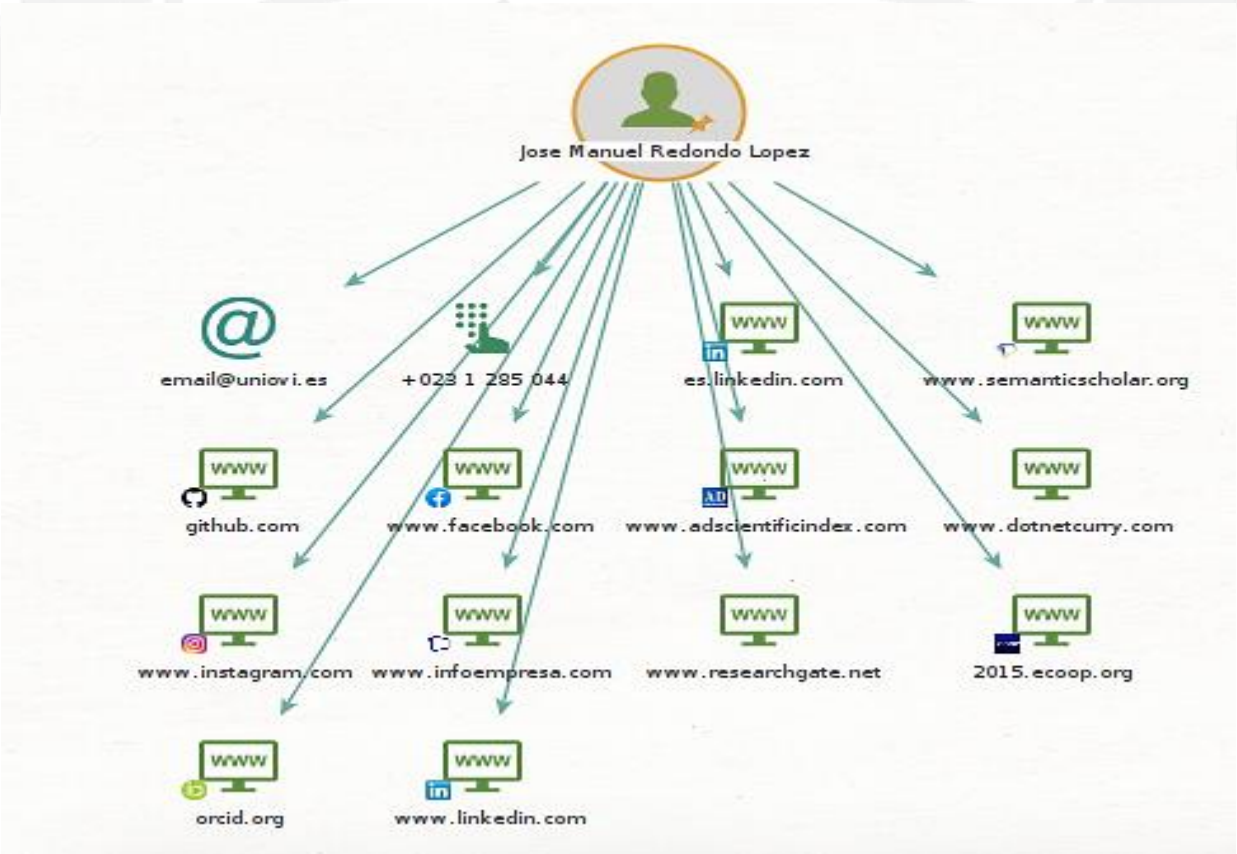


Figure 135. Information about me in *Maltego* (the phone number I have no idea where it comes from :D)

Searching on social networks

At this point, once obtained social media accounts of the investigated people, it is possible to use them to continue searching for information in an automated way as well. Normally to do that it is necessary to have an account of the social network that we want to investigate, since the **necessary API Key** to make the queries will only have it if we have an active account in it. The following image shows for example the result of using an entity that analyzes *Twitter*, something that will already be impossible to do for free because API usage will stop being free of charge from 09/02/2023[1].

On the other hand, there are *Transforms* associated with *Facebook* both in the palette and by third parties, and in the case of third parties there are also many other social networks (e.g.: *LinkedIn*, *Tiktok*...). Thanks to the use of these **third-party *Transforms*** (which we will see in the "***Transformation Center***" section), the power of the tool can be expanded.

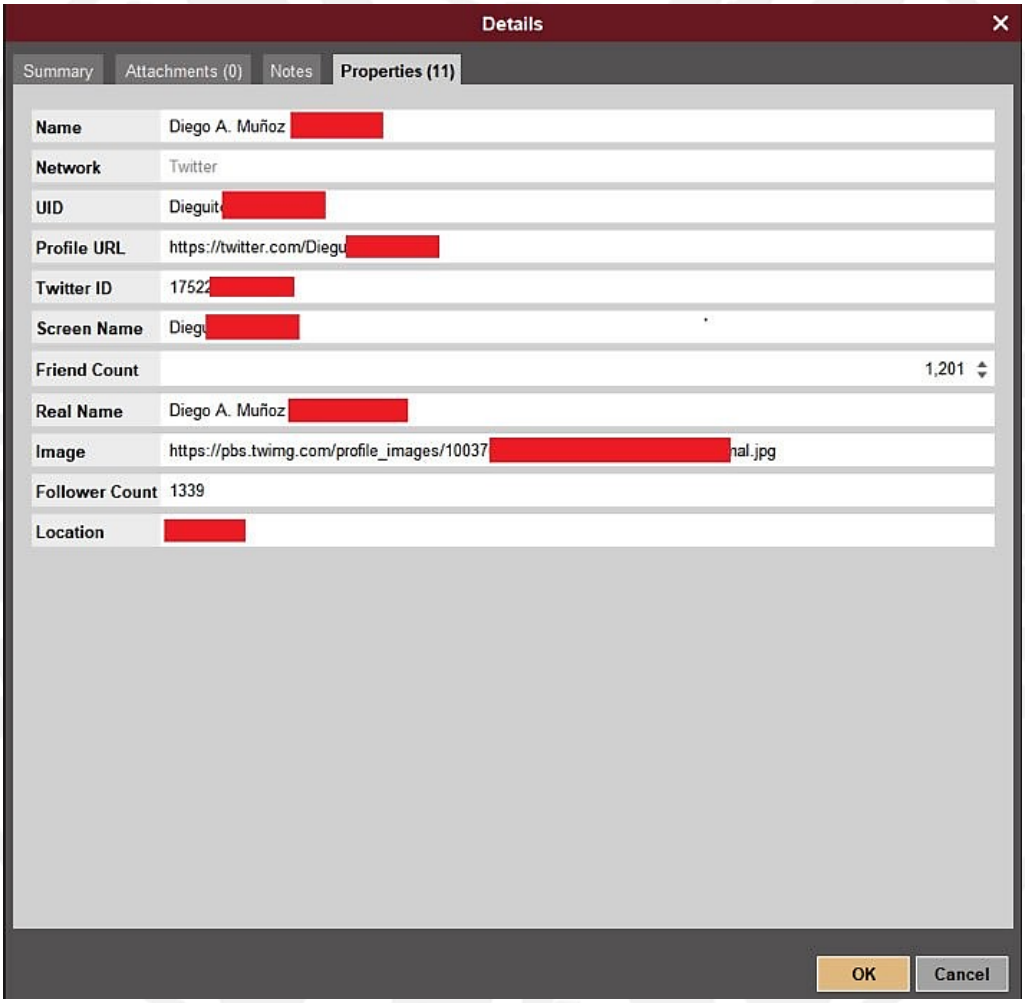


Figure 136. Profile of someone's Twitter account. Source: https://ciberpatrulla.com/maltego/

This information varies according to what is collected by *Maltego*. Sometimes you will find a lot of information and sometimes nothing, but that is the fun. No good research is easy or immediate, it needs work, cunning and perseverance to obtain good results, combine manual investigations with automatic recursive (refine), take notes, upload captures ... and take time. In this document, however, we are teaching

[1] I Heard a rumor that says that there exists a *GitHub* repo with valid *Twitter* API Keys tied to the social network phone apps (for example, the *iPhone* one), among others. If that is true, those API Keys may work with *Maltego*, as for the *Twitter* API they may be requests from a legitimate *iPhone* app and there is no reason to block them. But it is only a rumor, of course…

the bases for you to get an idea of the power of data mining associated with OSINT against entities that, as you can see, is a lot 😊.

Other Maltego entities

Maltego comes with many usable entities by default, and each of them can be dragged and used to search for different types of data. The number of entities available **depends on the version of the tool we use**. This image shows some of the CE version we have used so far.

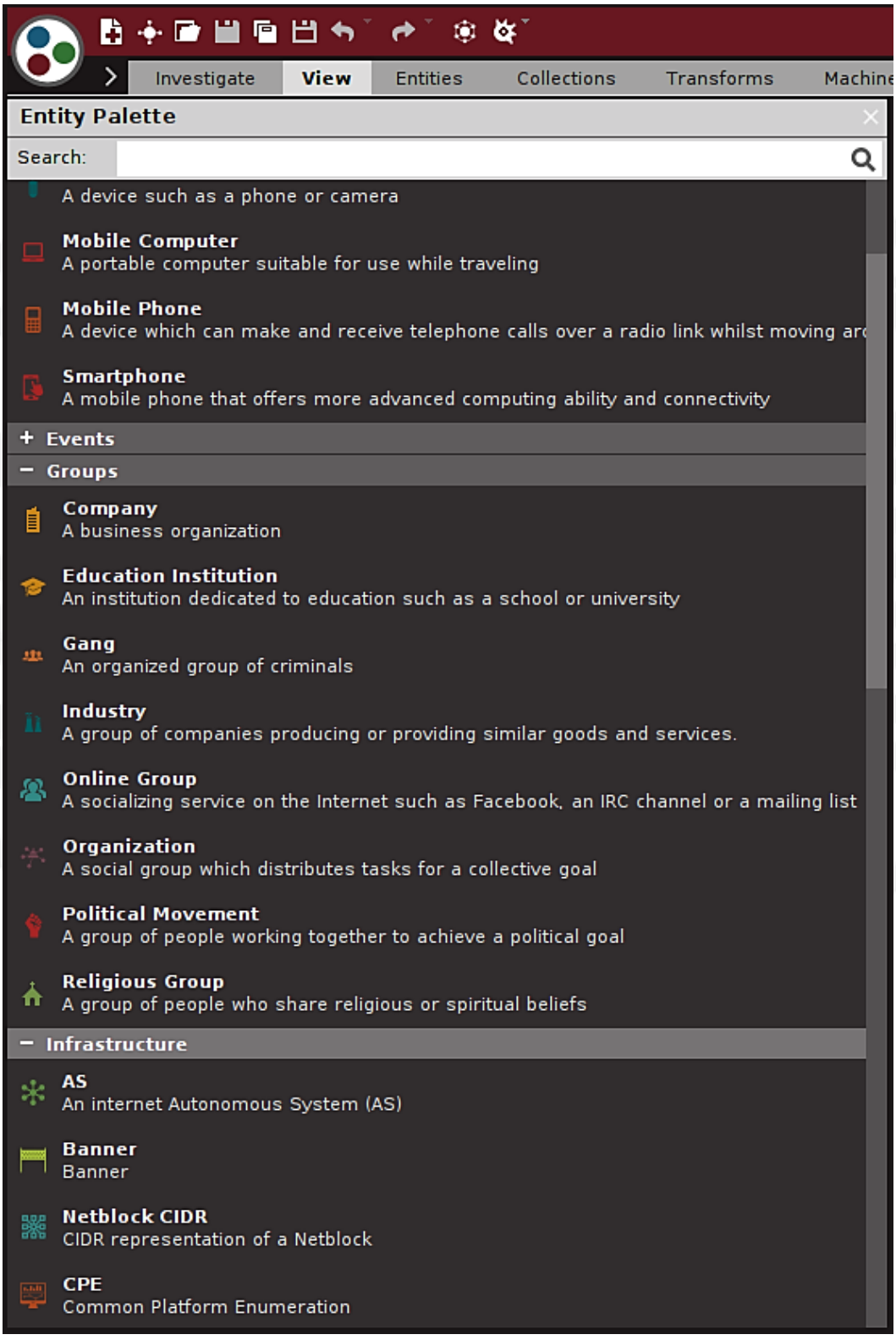


Figure 137. (Some) entities of the EC version of *Maltego*

In relation to the previous section, we also have several series related to known social networks, in this case those of *Twitter* are shown.

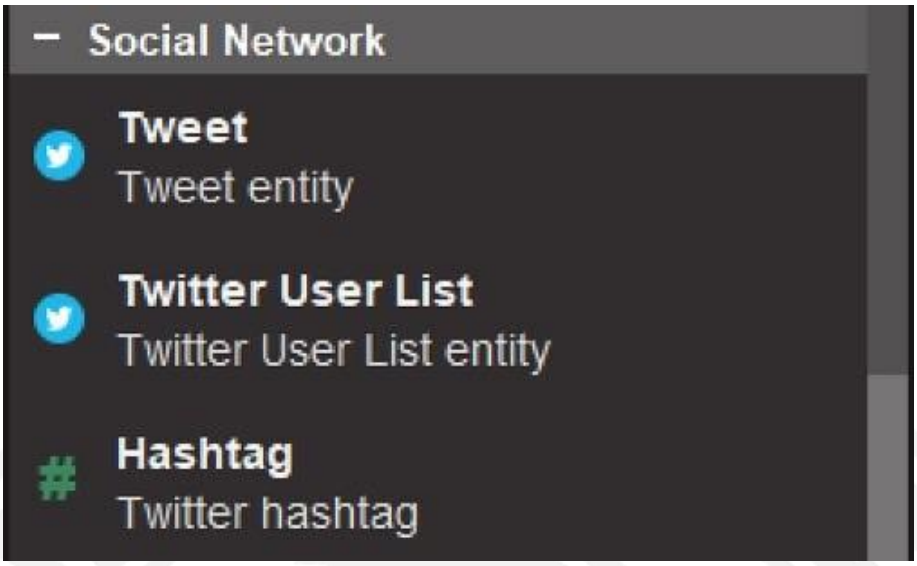


Figure 138. Twitter-related entities included in *Maltego*

As we can see, tools of this type facilitate certain types of information searches, since it can identify a large number of antecedents. However, perhaps in the most modern CE version entities are missing for certain social networks. But we can always find one made by third parties that can be imported. All of them can be controlled with the transformation center that we will see below.

The Transformation Center

The transformation center (next image) is what allows us to use external transformations, making the tool incredibly expandable and powerful with the developments of *Paterva* collaborators.

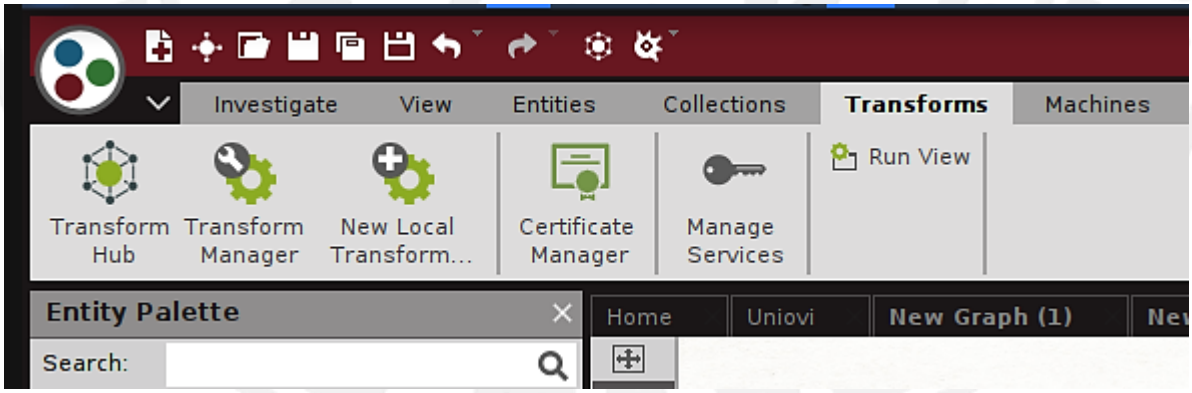


Figure 139. Maltego Transformation Centre

Once the "***Transform Hub***" is opened, we see a list of transformations available to install in our base *Maltego*, with their description to know if it will be useful or not. As we see, there are a lot of them available directly in the GUI of the tool.

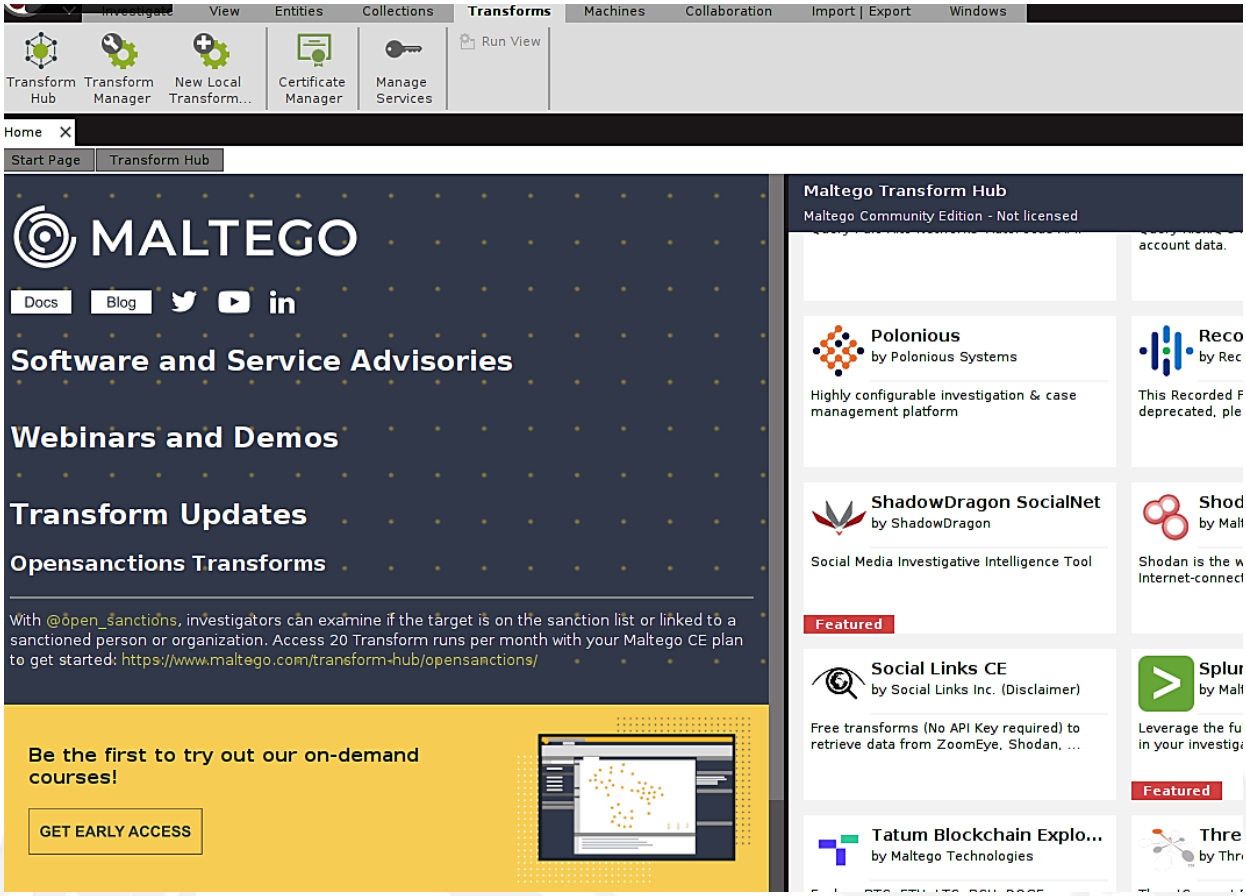


Figure 140. Transform Hub GUI of *Maltego*

This "*Transformation Center*" allows *Maltego* to access data from different tools or service providers using its official hub: https://www.maltego.com/transform-hub/. As we see, **we will not have all of them available**, since many are paid or for other versions of the tool. Luckily for us, we can filter by any criteria to stay only with those we can really access.

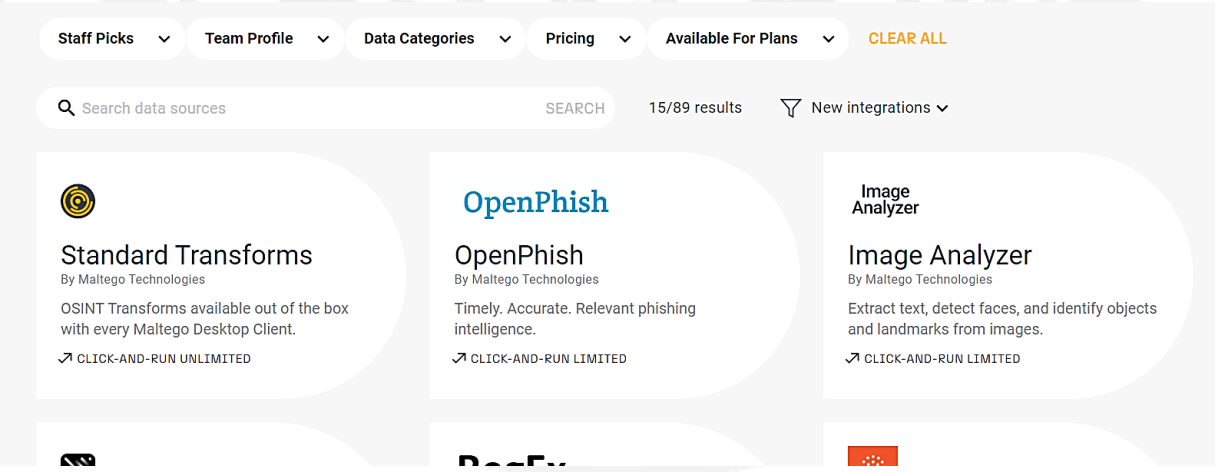


Figure 141. *Maltego* Transform Hub website

As an example, we will install the ***SocialLinks*** hub, which will give us some additional entities that we can use in any of the analysis already seen. To do this we simply look for the hub in the GUI and click on install, which as we see adds many more transformations, entities, and icons to enrich our graphics.

Figure 142. Installing *SocialLinks* for *Maltego*

Now, if we create a new *Graph* and go to the entities section, we can see that there is already the ***Social Links*** section, which allows us at this time and with the current version of the hub, to search for information using a "***Pipl***" entity (the people search network we saw before), a "***Company***" entity and background related to ***GitHub***. These entities are modified as the company that maintains it develops this module (*Maltego* (*Paterva*), only integrates them, does not control them). Again, to make use of some of them we need **API Keys** of the related services, such as this GitHub entity that asks us for the Key API, which we can obtain if we have an active GitHub account. As we can see, if we provide one, we can use it to extract information **from the members of a company or organization, their repositories, details of their code** and other details found on *GitHub* about the analyzed entity. Therefore, thanks to *Social Links* you can also cover another aspect of Internet search that we mentioned in **Seminar 1**, which until now we had not been able to cover 😊.

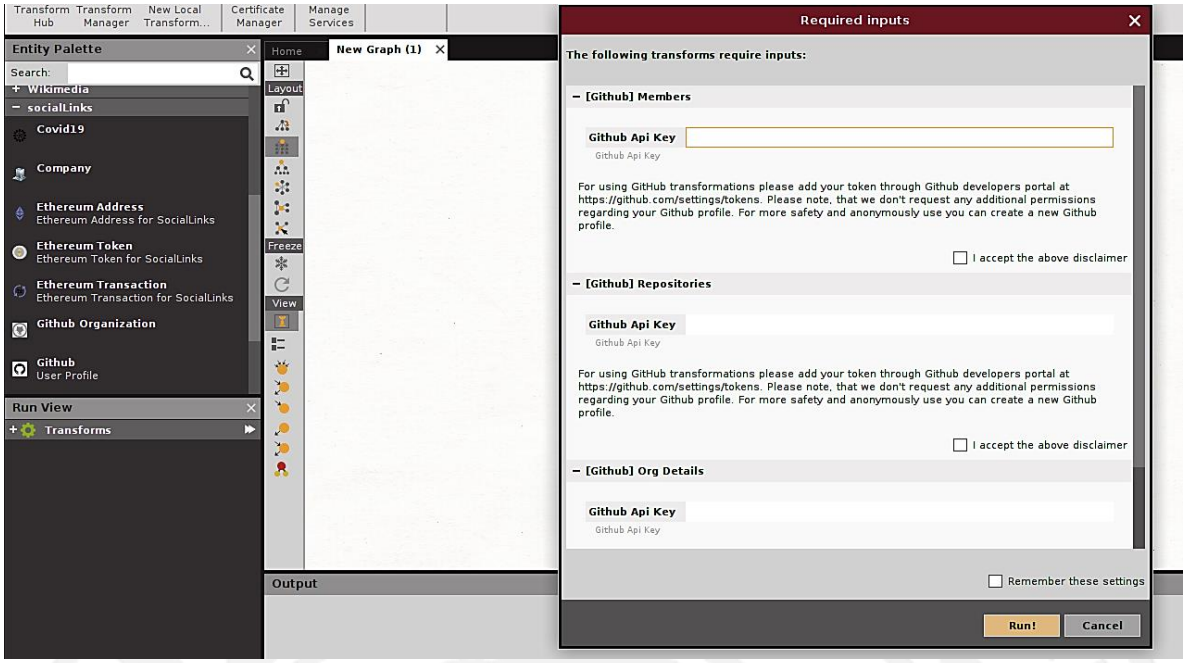


Figure 143. *Maltego SocialLinks GitHub* entity

By adding more entities to the *Maltego* palette, repeating previous searches **gives us more results**. For example, looking for myself again now not only pulls out websites associated with me, but supposed people and documents that are too. Interestingly, in this case **they are all false positives** (I do not know those people at all ...), although obviously it is not always the case. In the image we also see **how a document offers many properties that determine what it is without opening it**.

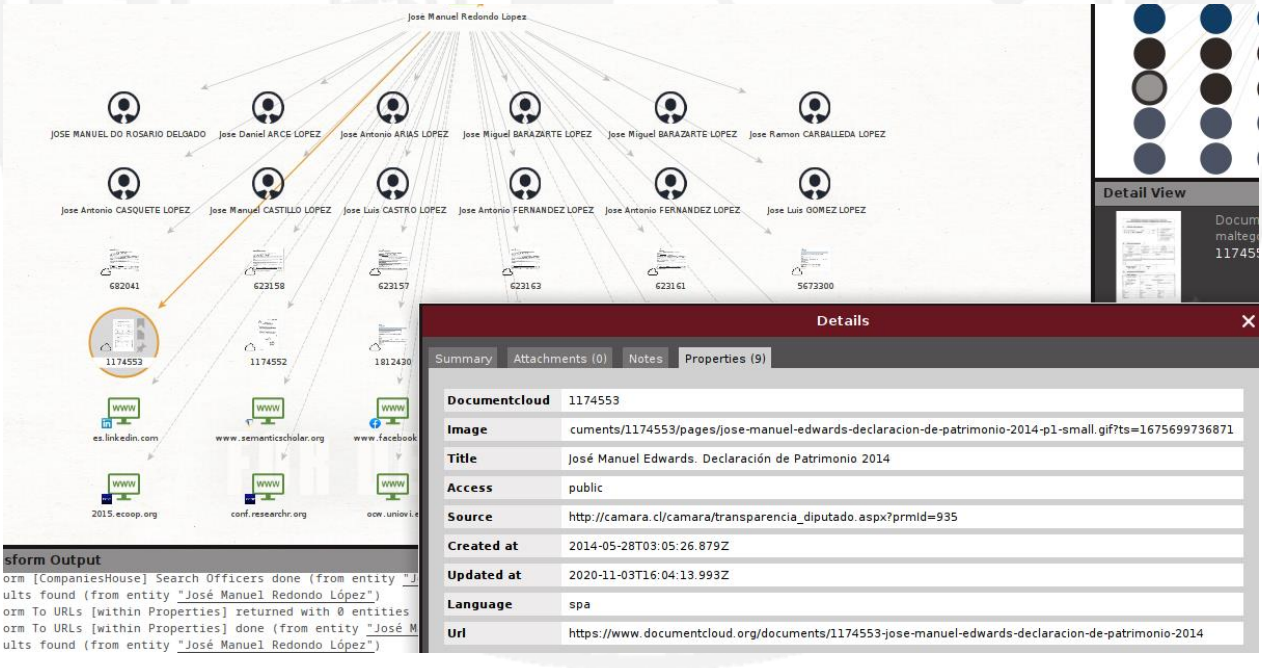


Figure 144. Search "enriched" with *Maltego SocialLinks* entities

To add new hubs, simply select the add option and follow the steps detailed.

Figure 145. Add new hub. Source: https://ciberpatrulla.com/maltego/

In the following example, we install the ***Shodan*** hub, and we have to enter the API that we got by creating our account as seen in **Seminar 1**.

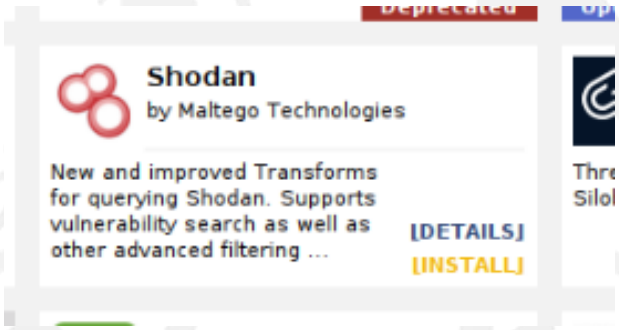


Figure 146. *Shodan* Hub for *Maltego*

Figure 147. Entering the Shodan API Key when prompted by the new hub. Source: https://ciberpatrulla.com/maltego/

Once this is done we could consult an **IP**-type entity and when we see its transformation options we will see that the one to do a *Shodan* scan has been added. **Realize what this means**: you can integrate the enormous amount of information that *Shodan* gives you about machines, with which *Maltego* gives you people (and machines too) on the other hand, with which **the "map" of your objective** will be much more complete. Now imagine that you follow this same philosophy with other services (which is what is done). You will see that this tool, in its paid version without limitations and in the hands of someone with experience, can be a real beast when working with open 😊 source data.

Figure 148. The power of *Shodan* within reach of a click (note that there are also *Transforms* from other hubs such as *Censys* and other search engines 😉)

In this GitHub *repository* there are linked a **large number of transformations** for *Maltego* that can be useful in different scenarios, some of them interacting with engines and frameworks that are also seen in this document: https://github.com/cipher387/maltego-transforms-list/

Now you only have to continue interacting with *Maltego* and test each of its entities and functionalities, integrate new APIs for specific searches and develop our own methodology for our research. To the moon! :D. Obviously with the CE version you will be limited, but even so, the power you will have is much greater than not using it 😉.

Export the information

Maltego has the option of exporting its results in order to include them in documentation and/or present them to a client who has hired us to do an audit. To do this it is only necessary to go to the menu and select the tab "*Import | Export*" to finally use "*Generate Report*"

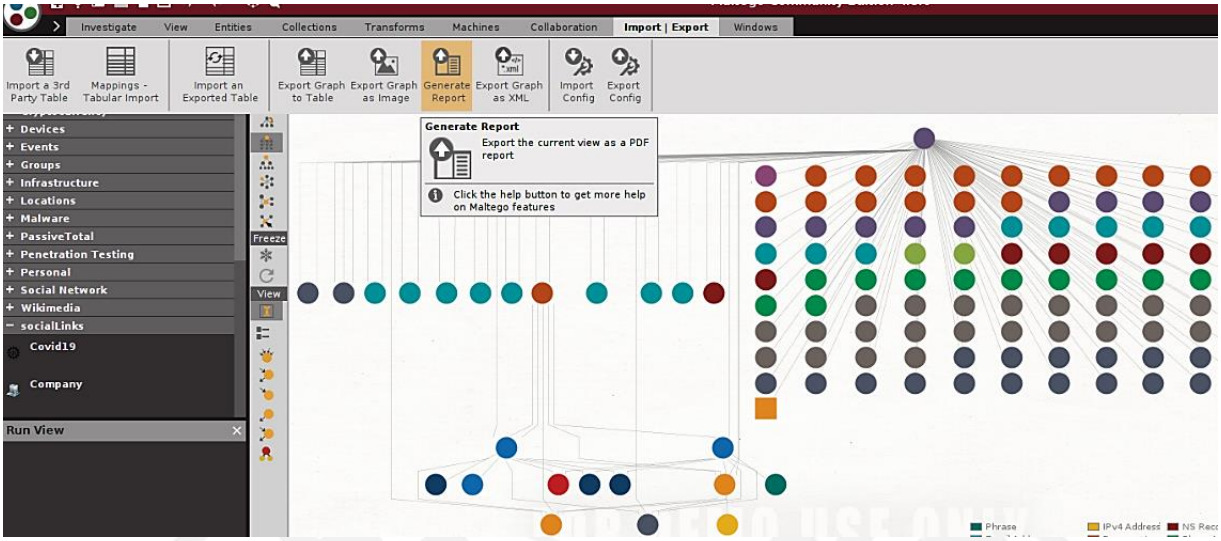


Figure 149. Reporting in *Maltego*

We generate a report and save the results in PDF.

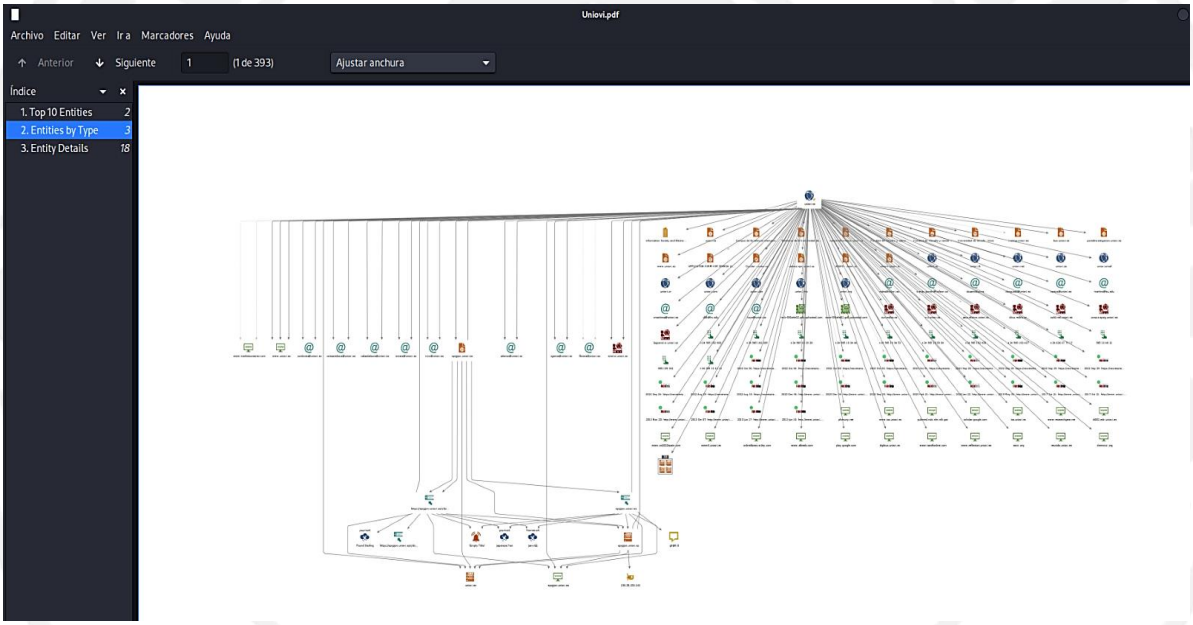


Figure 150. *Maltego* PDF report

We also have the option to import as a table or image, as you need or like. Keep in mind that **this report can be very extensive**. It is not limited to showing only the graph, but tables of data and information about each entity found and analyzed exhaustively. This is because it is the type of information needed by the people who analyze it. To give you an idea, only the report resulting from analyzing `uniovi.es` and `epigijon.uniovi.es` and their entities that we saw at the beginning of this section of automatic OSINT almost reaches 400 pages. The good thing is that there is nothing left undocumented 😊.

I want to know more!

In this document we have only seen a typical common use of *Maltego* to perform OSINT research oriented to this course. But the tool offers much more 😊. One of the challenges that can be more complex at the beginning is how to do a PoI (***Person of Interest***) research in a coherent and orderly way, given the large amount of information that we can obtain as soon as we expand the entities and make transformations. This diagram can help you with this:

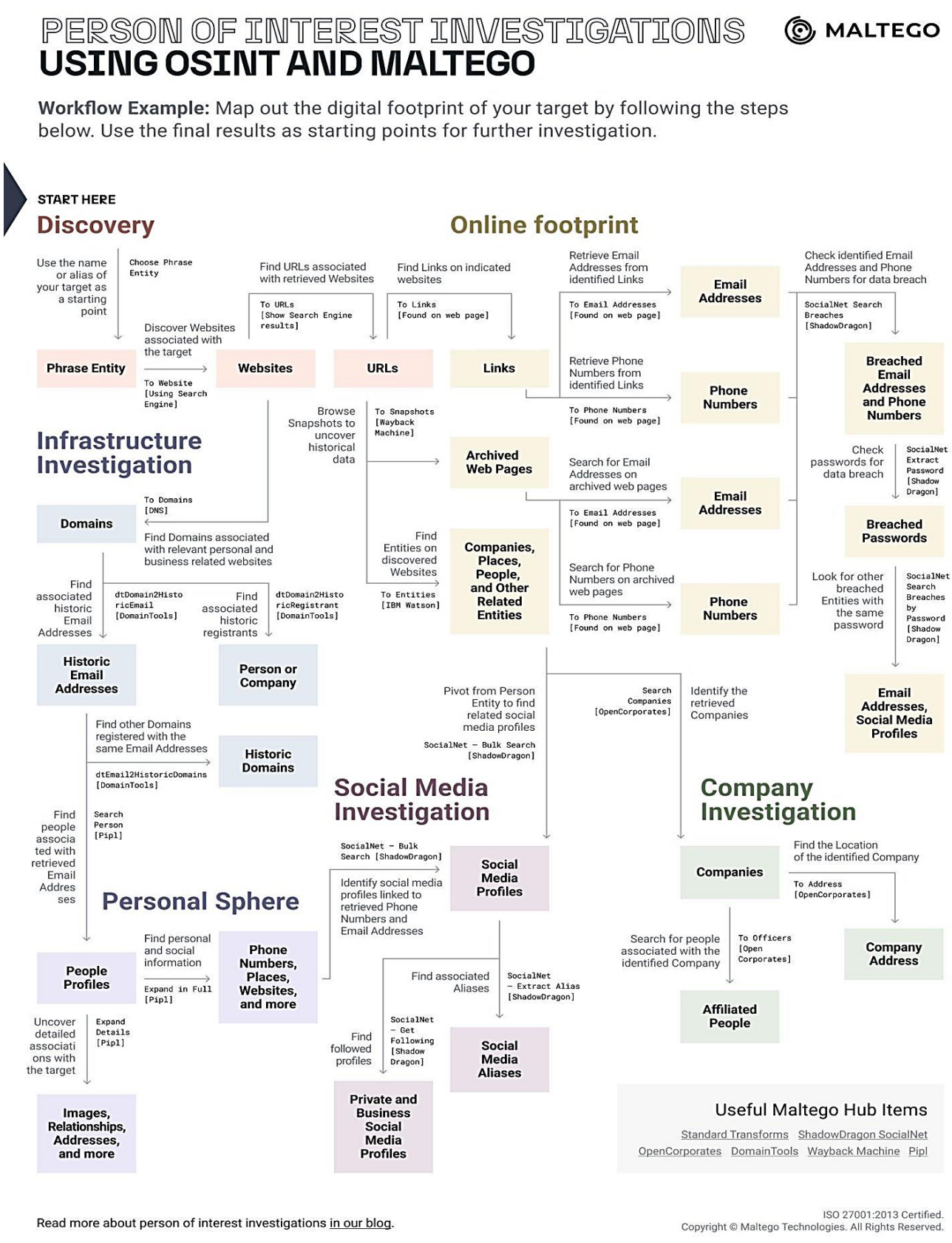


Figure 151. Using the power of Maltego correctly involves planning, this scheme can help you with it

If you want to learn more about *Maltego*, you can check these links:

- **Official *Maltego* website**: https://www.paterva.com/web7/
- **Documentation**: https://www.paterva.com/web7/docs.php
- ***Maltego* User Guide** (Version 3): https://www.paterva.com/web7/docs/M3GuideGUI.pdf

- ***Paterva* YouTube channel**: https://www.youtube.com/user/PatervaMaltego
- ***Cylon Null* YouTube channel**: https://www.youtube.com/channel/UCseLiPDQKBrKJfqtiHuH07w/videos

"Automatic" OSINT: Search Engines and Threat Intelligence

Introduction

Cyber threat intelligence (CTI) is information based on knowledge, skills and experience about cyber and physical threats and actors who make use of those threats. It includes both determining their existence in a given context and assessing the severity of these. It aims to help mitigate potential attacks and harmful events occurring in cyberspace. We talk about this term here because sources of threat intelligence of this type include:

- Information obtained via **OSINT**
- Information obtained from **social networks**
- Information obtained from **human actors**
- **Technical information**: device log files, forensically acquired data or Internet traffic, data obtained or derived from what is obtained from the *Deep Web* or the *Dark Web*...

That is, this document covers a good part of these sources 😊. If being a "cibermaruja" does not convince you, you can say that you are "obtaining threat intelligence". 😉 Of course, to say that you obviously have to study actors (humans or machines) **that may be threats to the infrastructures or clients that you protect or audit**, not anyone on the Internet (do not be a cheater 😉).

In recent years, threat intelligence has become a crucial part of companies' cybersecurity strategy, as it allows companies to be more proactive in their approach and determine which threats pose the greatest risks to a business. **This causes companies to actively look for their vulnerabilities** and prevents attacks before they occur. Since this method is gaining importance nowadays, knowing techniques like the ones I have written gives you a plus work 😊.

So far we have seen ways to get threat intelligence by manual means and with tools that automate part of them. But in this section we are going to present **search engines**, mostly free, which allow us to expand the information of the objectives **passively** ("do not engage") and complete our study.

The objective of these engines is multiple:

- You can use them to make a **quick evaluation** at no cost and anonymous of a study objective. In that case, I recommend using several. Not only because they complement each other, but because a target could block some of them. For example, it is typical for a target to block *Shodan* because it is the most popular. But then, however, the same information is found in others who do very similar things but are less known.
- You can **populate your *Maltego* chart** with the data you get from several of them. Moreover, surely *Maltego* has a module that directly allows you to use them and that this is done automatically. Hopefully, the module will be free and compatible with the CE 😉 version.
- If you are from the "defender" point of view, **you can clearly see what "bad guys" may know about you** by looking at the Internet. This is just as it sounds. This information is there, available to anyone

who knows how to look for it and can be used against you. So, you better "cover" the "hole" before an “unpleasant someone” sneaks in.

If your company is large and important, it is very convenient to consider paying for a subscription to some of these search engines and monitor their results to find threats that affect it. No joke, it is something preventive and very effective.

Next we will see a set of search engines for CTI well known and used worldwide, classified by the type of information they are able to find primarily.

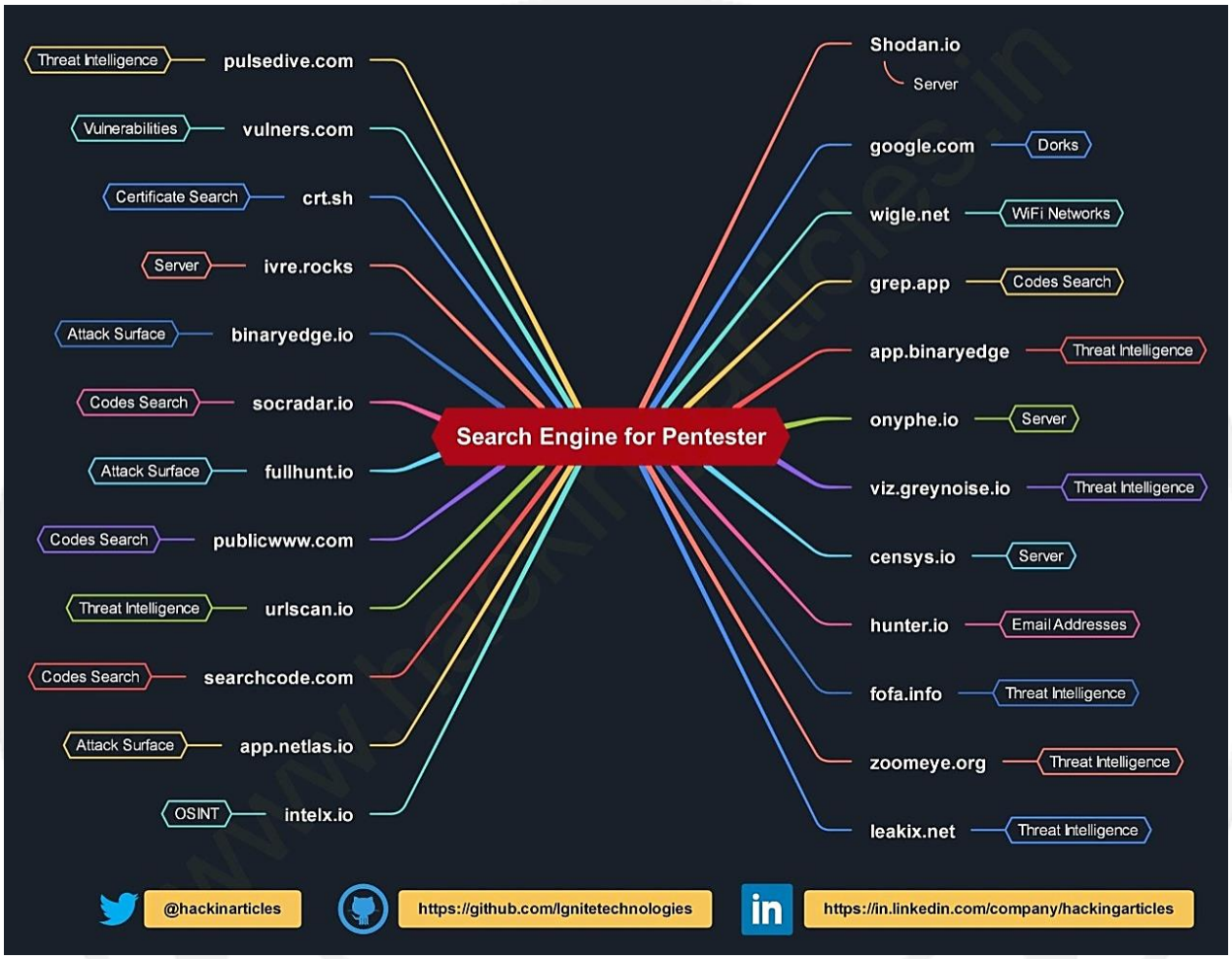


Figure 152. General outline of useful engines for different purposes and searched information

OSINT APIs: OSINT in our programs!

OSINT is not just a technique that is done manually or automatically with specific tools that discover information about different data sources. **There are also** APIs that allow us to find diverse information about a multitude of entities. As APIs, **they allow us to design programs that make use of this information** for some of our research needs, combine them, or make inferences by program from them. Indeed, as software engineers, we can now create programs that use OSINT data! 😊

APIs usually require an API Key to use them, and these are no exception. Some offer it for free with a limited request/time ratio (enough for sporadic testing). Others are not free, so they only make sense if we dedicate ourselves professionally to this. We must also ensure that APIs include the countries we are interested in, as it is common to find APIs that **only offer data from United States**, whose data processing and publication policy is very different from that of, for example, Europe. Before paying, we must be sure that the results will be useful for us!

A site that provides access to different APIs is this one: https://inteltechniques.com/tools/API.html

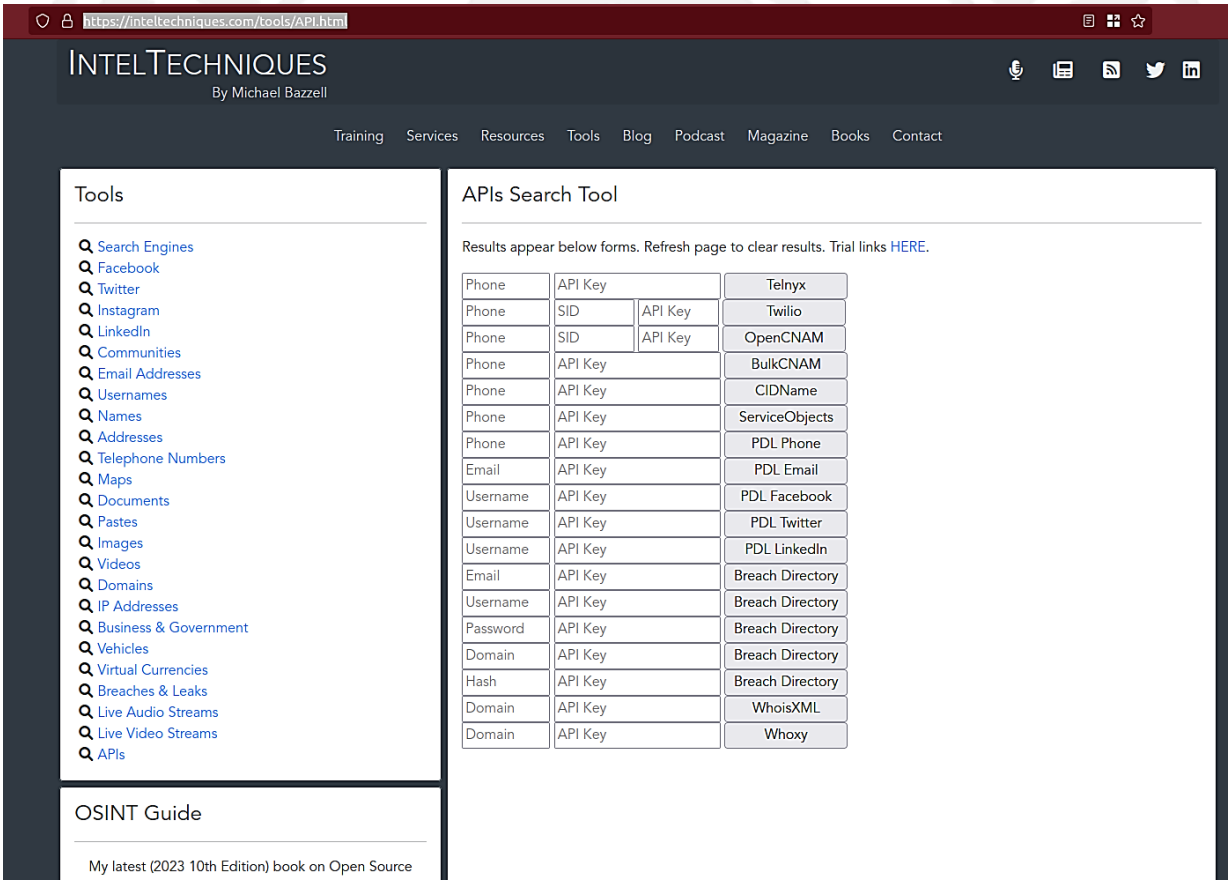


Figure 153. OSINT APIs from *Intel Techniques*

We also have the following *GitHub* repository with a lot of references to APIs to get any kind of data we can think of: https://github.com/cipher387/API-s-for-OSINT

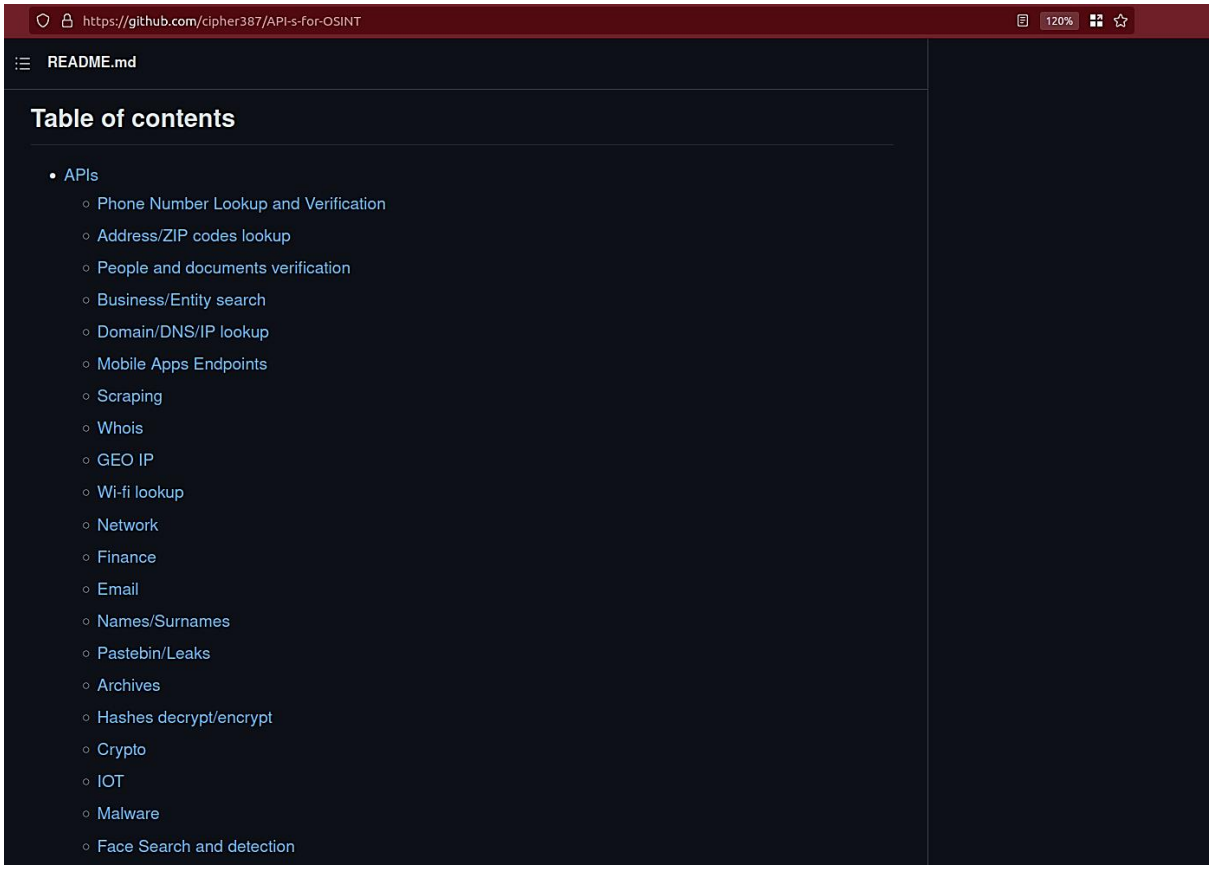


Figure 154. GitHub repository with OSINT APIs

Interactive OSINT tool trees classified by performed task

With the number of existing OSINT tools, it is often very difficult to know which tool to use in each case, either because we do not know about one, or simply because we do not remember it. This type of tools that represent an interactive tree of OSINT tools can be very helpful for this, and for that reason in this section we present some examples

Malfrat's OSINT (https://map.malfrats.industries/)

Map that gives access to different tools depending on what you are looking for in the drop-down tree that is shown. As seen in the image, it covers a substantial number of applications and usages.

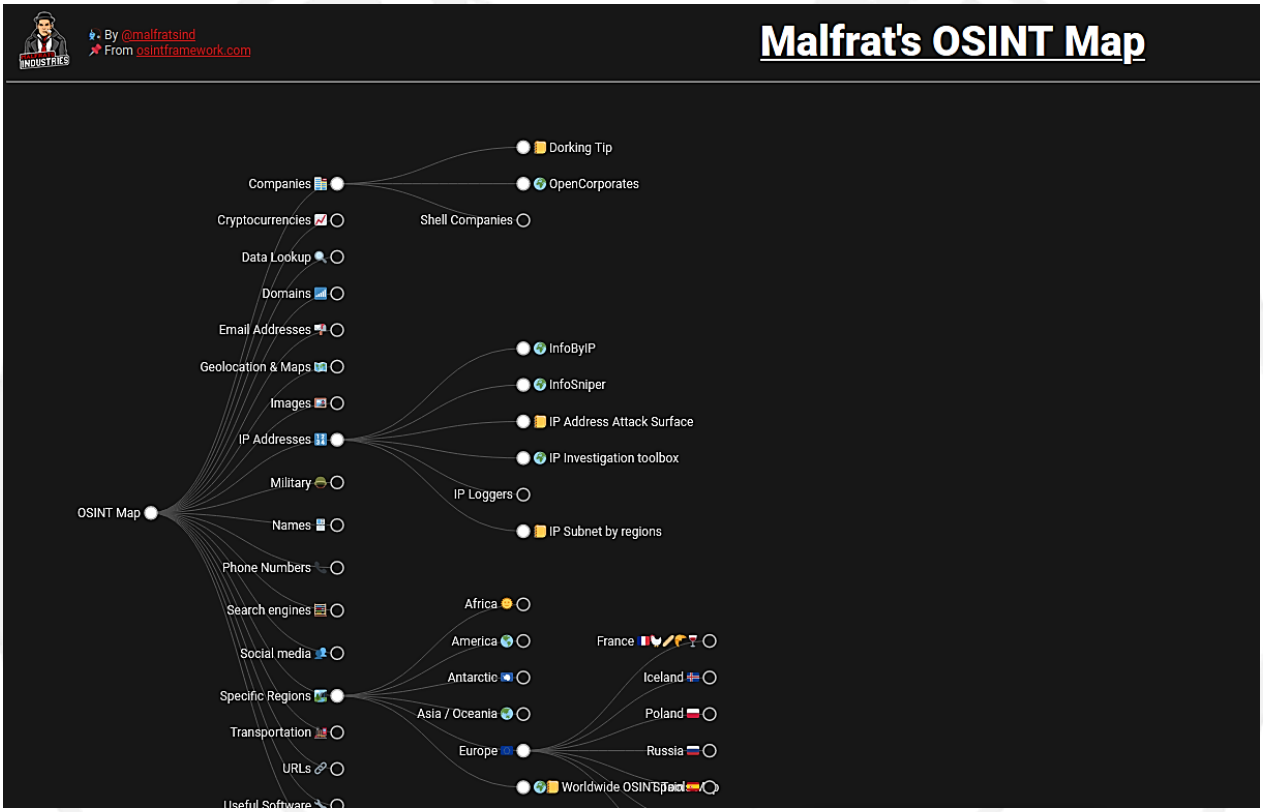


Figure 155. Malfrat's OSINT

OSINT Framework (https://osintframework.com/)

Another map that gives access to different tools depending on what you are looking for in the drop-down tree displayed but focusing more on information about people. *Do you need to find something specific about a person?* Here are the tools you can use that are best suited for this type of information. It is a perfect guide for those who do not know exactly how to proceed when it comes to finding information.

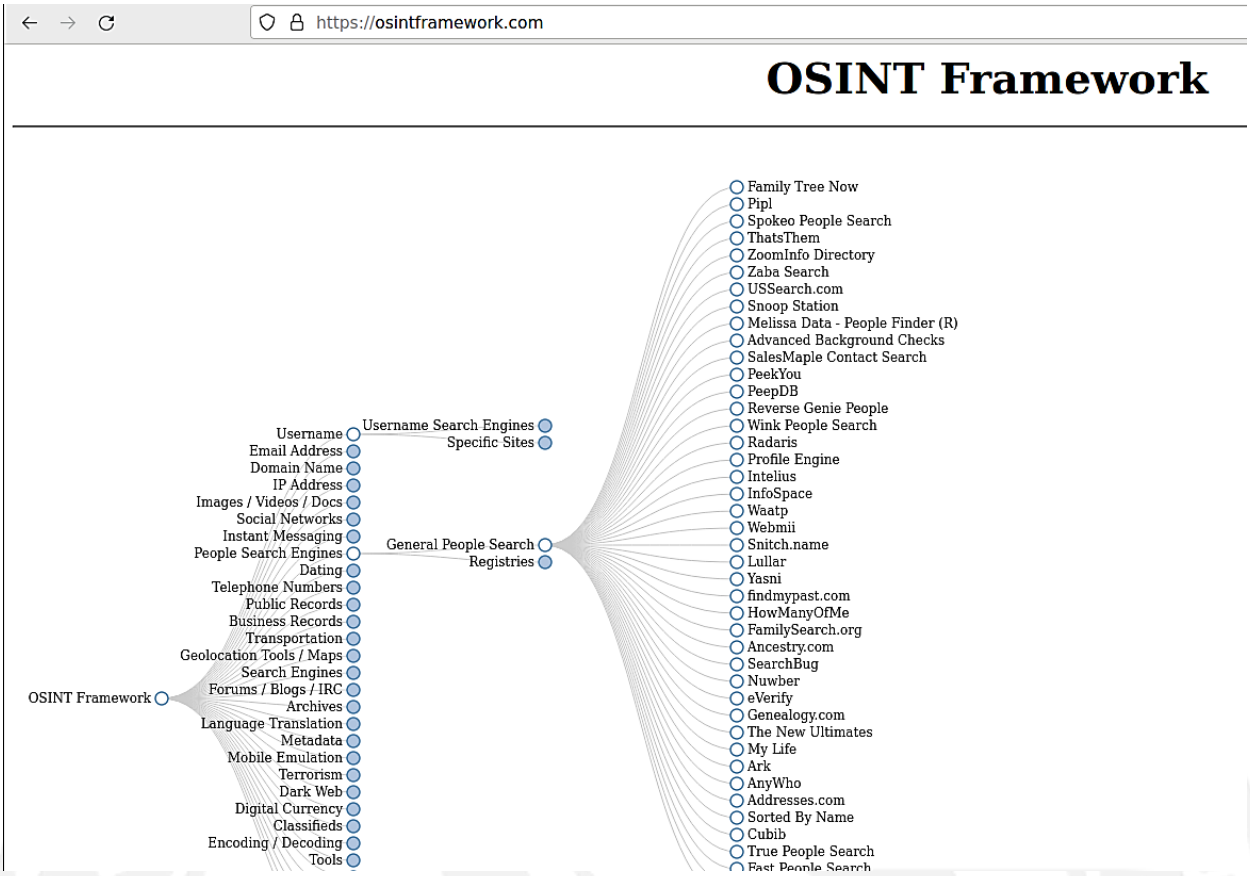


Figure 156. OSINT Framework

Search engines for machines and networks

There are a lot of search engines out there for internet research, and in this section we are going to look at several of them. However, they **are not all of them,** and in this image you can find more than the ones included in this section and in the document in general. If you think one of them might be good for you, do not hesitate to use their services!

Figure 157. Search engines for researchers are many, and each one has its own purpose

Binary Edge (https://www.binaryedge.io/)

It collects data from the internet to feed its own search engine much like *Shodan*, *Censys*, and others we have seen or will see. It allows up to 250 queries per month free and from 10 dollars per month for 5000 consultations. You can find many **examples of useful queries** at: https://github.com/balgan/binaryedge-cheatsheet

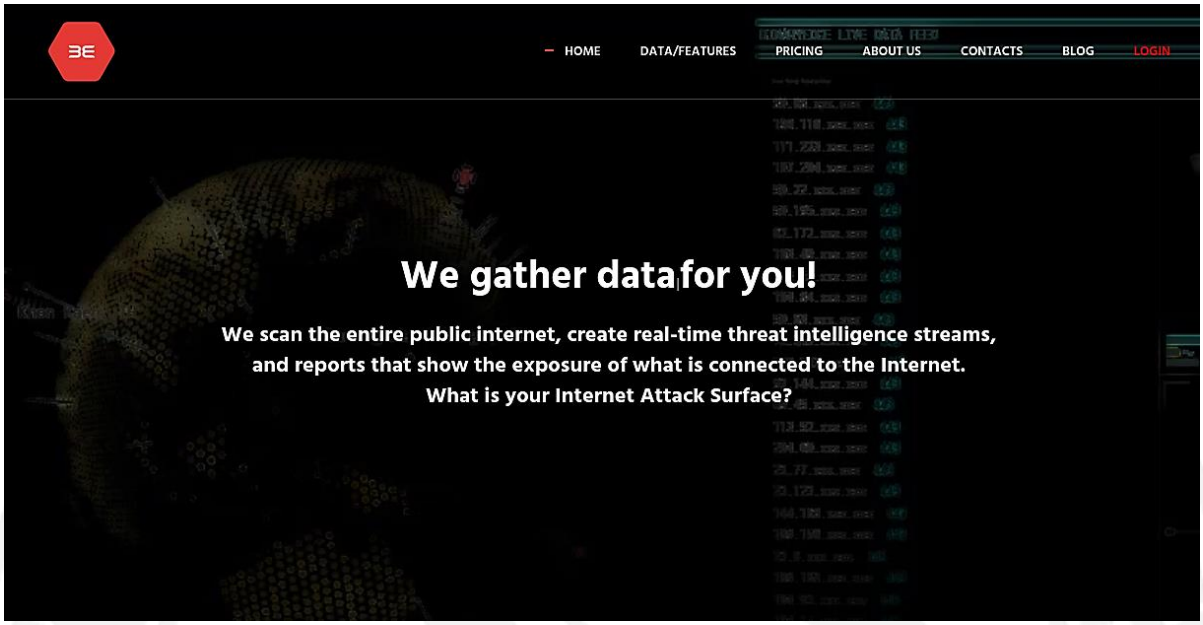


Figure 158. ***BinaryEdge***

Censys (https://search.censys.io/)

It is the engine that we saw in **Seminar 1** and that assesses the attack surface for devices connected to the Internet. As we saw, this search engine finds and detects versions of web servers, database servers, unpatched vulnerabilities, routers, operating system versions, insecure services, web application firewalls... and, in general, a lot of information that allows you to recognize any site to see its security problems.

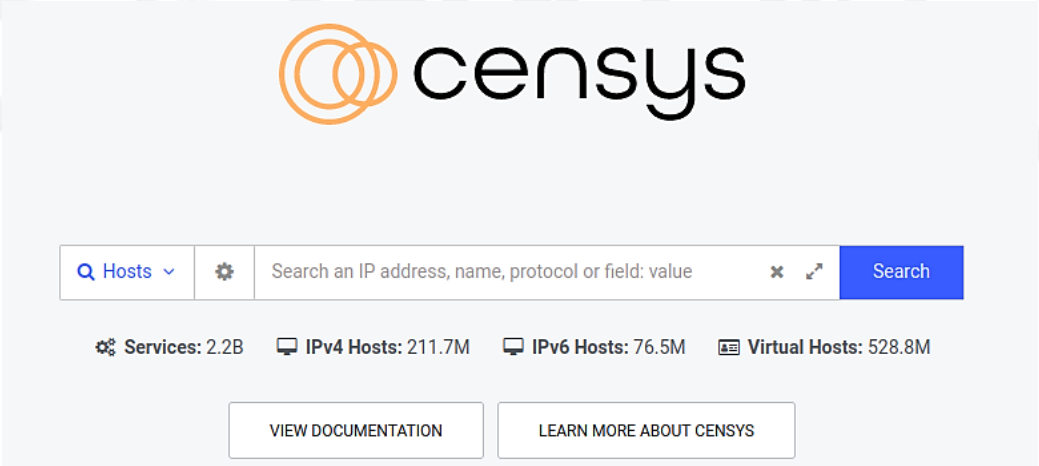


Figure 159. ***Censys***

You can find many **example queries** to work with it here: https://cheatography.com/himajedi/cheat-sheets/censys-search/. The following cheatsheet also contains **frequent queries** you can do in *Censys*.

CENSYS Quick Start Reference — censys.io

What is Censys?

Censys is a publicly available search engine, similar to Shodan but unique in its own right, which scans the entire Internet for a limited number of services and enumerates discovered services by their banner responses, indexes that data and makes it searchable.

Censys stores the information in structured fields which can be queried specifically for enumerating data on hosts, services and (in particular) web certificates.

Be sure to use the **'Raw Data'** option on any discovered host to see all of the data types Censys has stored.

Censys also indexes WHOIS data which can be viewed from the same menu under **"Raw WHOIS"**.

IP Addresses & Subnets

Single IP Address – Search for findings on single IP
Example: **52.179.197.205** *or* **ip:52.179.197.205**

IP Subnet by CIDR – Search across a specific CIDR
Example: **ip:52.179.197.0/24**

IP Subnet by Range – Search across a specific range
Example: **ip:[216.189.94.1 TO 216.189.94.32]**

Hostname – Search on result of a DNS "A" / host entry
Example: **a:panerabread.com**

Mail Servers – Search on DNS "MX" entries for domain
Example: **mx:panerabread.com**

Port – Find any instances of active services on a port
Example: **ports:21**

Service – Search for instances of specific services
Example: **protocols:"21/ftp"**

Autonomous System Number (ASN) – Search by ASN
Example: **autonomous_system.asn: 7018**

Physical Location

Country – Search by country code
Example: **location.country_code:"US"**

City – Search by city name
Example: **location.city:Paris**

State – Search by state name
Example: **location.province:South Carolina**

Zip Code – Search by postal ZIP code
Example: **location.postal_code:92127**

Geo : Latitude Range – Search GPS coordinates - Latitude
Example: **location.latitude:[45.0 TO 59.0]**

Geo : Longitude Range – Search GPS coordinates - Longitude
Example: **location.longitude:[15.0 TO 18.5]**

Operating Systems & Products

Operating System – Search by operating system type
Examples: **metadata.os:Windows**

Product (Web Service) – Search by known product name
Example: **443.https.get.metadata.product:nginx**

Manufacturer – Search for known manufacturers
Example: **metadata.manufacturer:"Huawei"**

Microsoft SMBv1 – Search for instances of SMBv1
Example: **445.smb.banner.smbv1_support:true**

Dates & Ranges

Date: After – Search for findings that appear after a date
Example: **updated_at:[2018-12-15 TO *]**

Date: Before – Search for findings that appear before a date
Example: **updated_at:[* TO 2018-12-31]**

Date : Range – Search for findings that appear within a range
Example: **updated_at:[2018-12-15 TO 2018-12-31]**

Web Apps

Page's Title – Search for text in page's title
Example: **443.https.get.title:"Index of /ftp"**

Page's HTML Body – Search body of webpage for text string
Example: **443.https.get.body:"XML-RPC server accepts"**

Web Technologies – Search for specific web technologies
Example: **443.https.get.metadata.product: php**

TLS Version – Determine most recent version supported
Example: **443.https.tls.version:TLSv1.2**

SSLv3 – Find instances of SSLv3
Example: **443.https.ssl_3.support:true**

Expired Certificates – Search for expired HTTPS certs
Example:
443.https.tls.certificate.parsed.validity.end:[2018-12-31 TO *]

Self-Signed Certificates – Search for expired HTTPS certs
Example:
443.https.tls.certificate.parsed.signature.self_signed:true

Invalid Cert Signatures – Find invalid cert signatures
Example:
443.https.tls.certificate.parsed.signature.valid:false

Trusted Certs – Determine trusted certs by browsers
Example: **443.https.tls.validation.browser_trusted**

Heartbleed – Find potential instances of Heartbleed vuln
Example:
443.https.heartbleed.heartbleed_vulnerable:true

Tags

A list of common tags that I've found useful:
bacnet, database, DSL/cable modem, embedded, Heartbleed, industrial control system, known-private-key, modbus, mssql, mysql, network, oracle, postgres, printer, rdp, remote_display, raspberry pi, scada, smb, vnc

A special thank you to the Censys team (@censysio) and the University of Michigan (@UMich) !!!

By Michael Holcomb (@mdholcomb)

Figure 160. Cheatsheet with Censys frequent queries

DNS: Specific tools for servers and domains

MX ToolBox (https://mxtoolbox.com/)

If our target has their own domain, this website can offer us a lot of public information about it that we can use to continue looking for information about him. It also serves for domain diagnostics, not just searches.

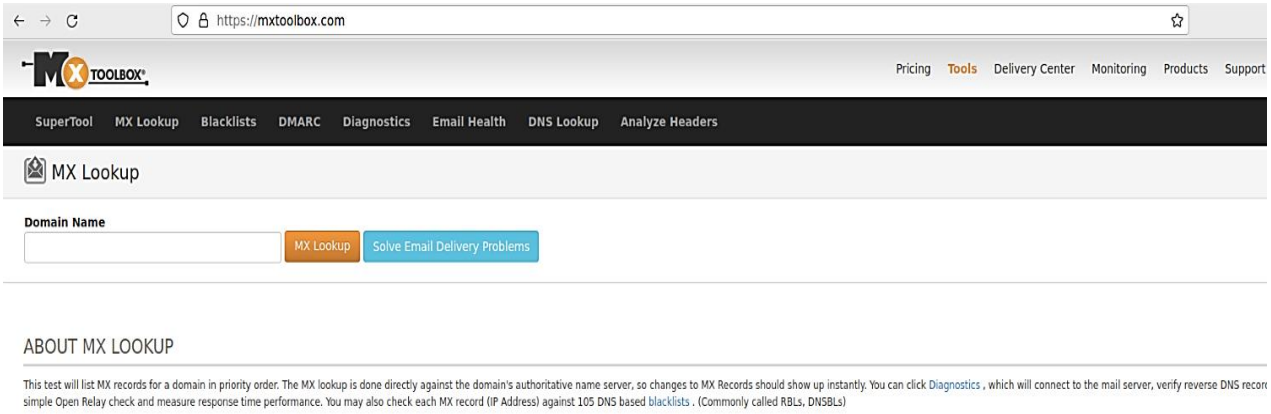


Figure 161. *MX Toolbox*

Domain investigation Toolbox (https://cipher387.github.io/domain_investigation_toolbox/ip.html)

This page offers us information about IPs and DNS by aggregating the information offered by various third-party tools, as seen in the image. It is something like a kind of information aggregator that allows you to know a lot of things about a target in a single operation.

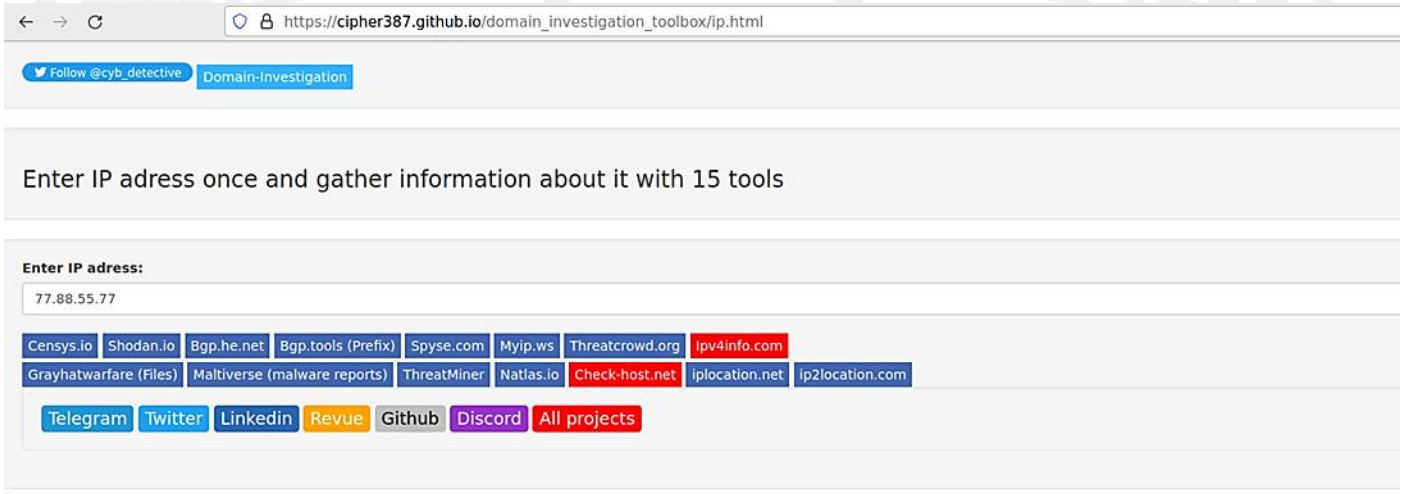


Figure 162. *Domain Investigation Toolbox*

Web tool with individual operations: View DNS (https://viewdns.info/)

Sometimes what we need is not a massive amount of information about a DNS name, but rather a set of specific data. While all this can be done by command line tools, this website allows you to do it online, separating each operation from the others as seen in the image.

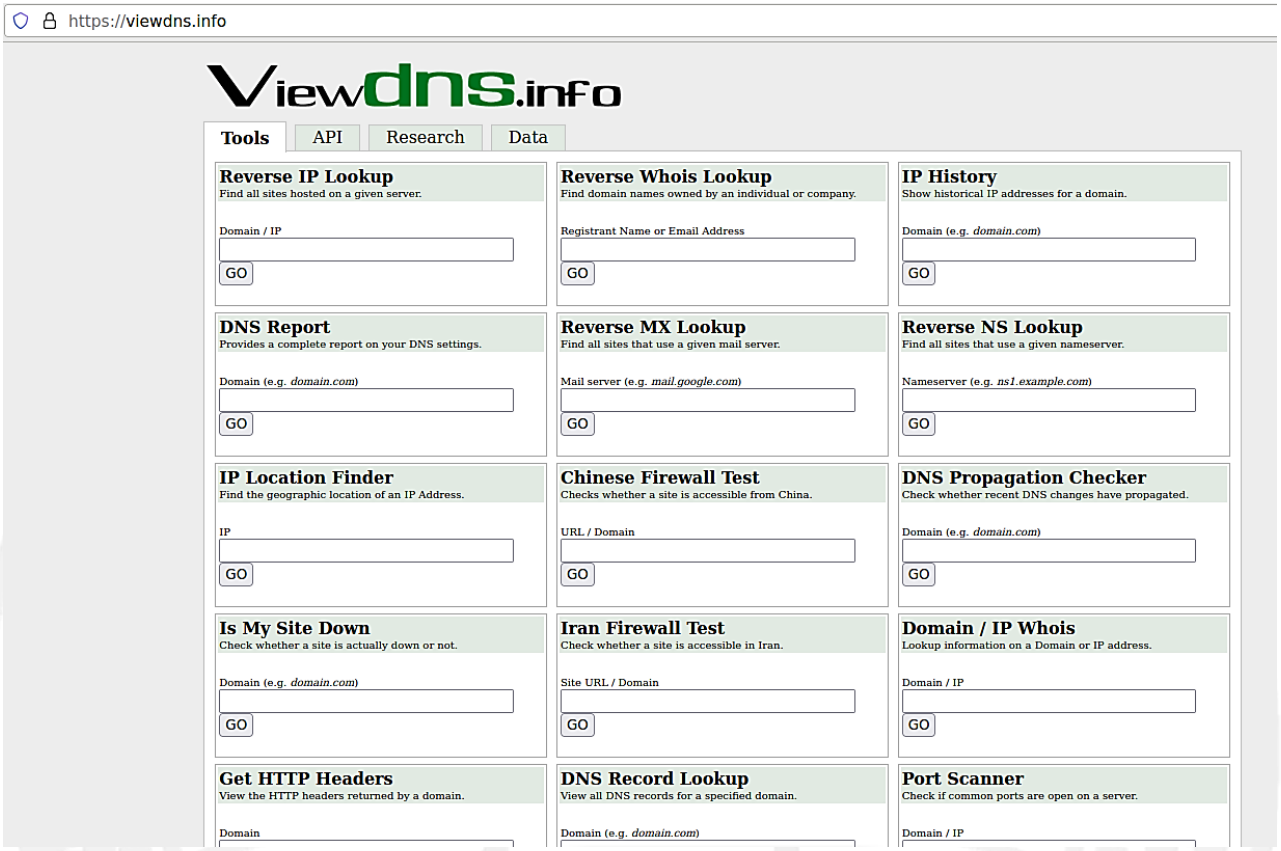


Figure 163. *View DNS* with a collection of individual DNS operations

FOFA (https://en.fofa.info/)

A search engine for Threat Intelligence, which also allows you to manage the attack surface of a company more easily. You can find a **very useful guide to queries** with this engine here: https://github.com/FofaInfo/Awesome-FOFA/blob/main/Get%20Started%20with%20FOFA/A%20Beginner%E2%80%98s%20Guide.md

Figure 164. FOFA

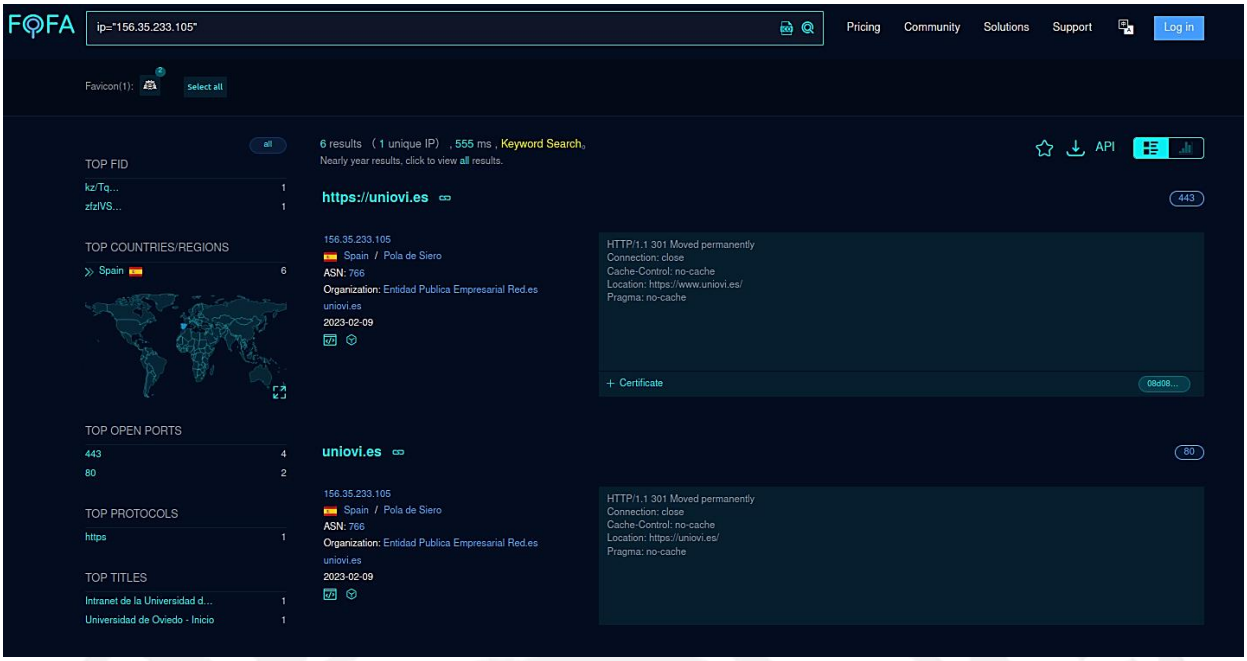


Figure 165. Example of FOFA search result on Uniovi

FullHunt (https://fullhunt.io/)

It is a database of attack surfaces from all over the Internet. It allows you to discover attack surfaces, monitor them for exposure, and continuously scan them for the latest security vulnerabilities.

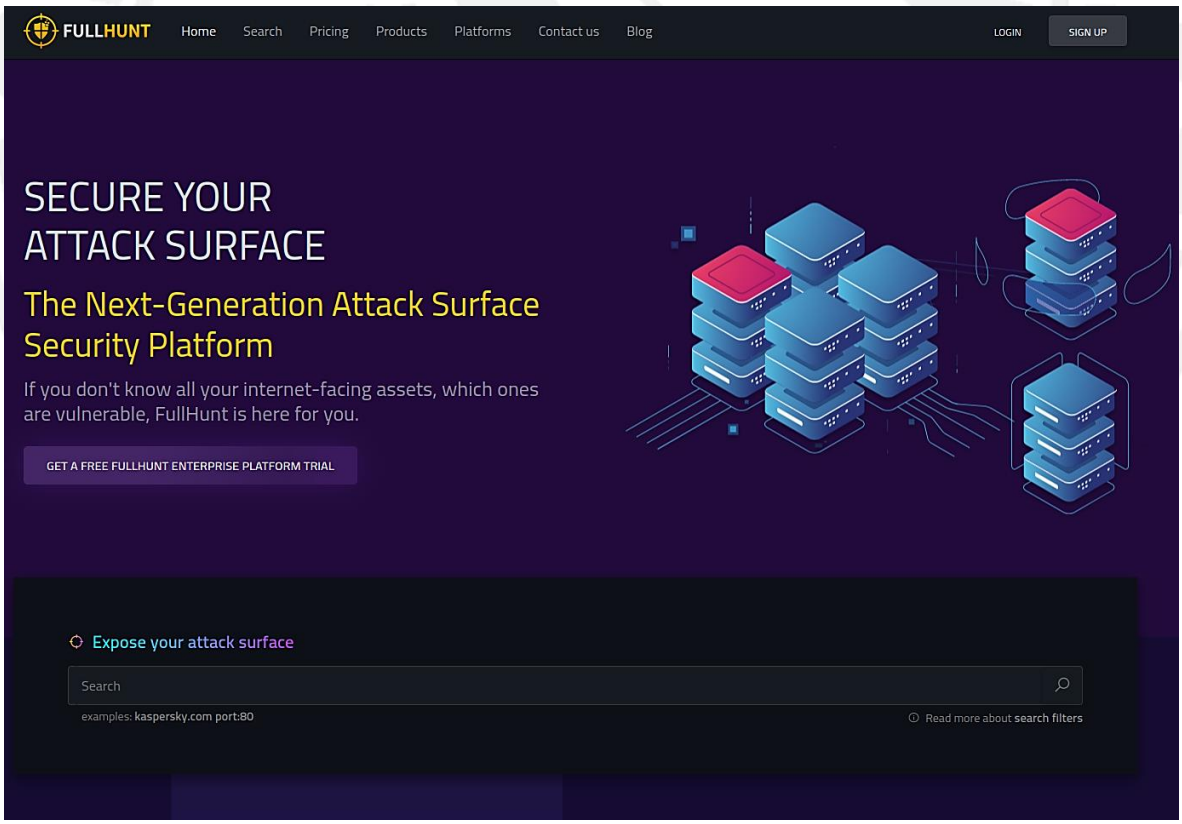


Figure 166. *FullHunt*

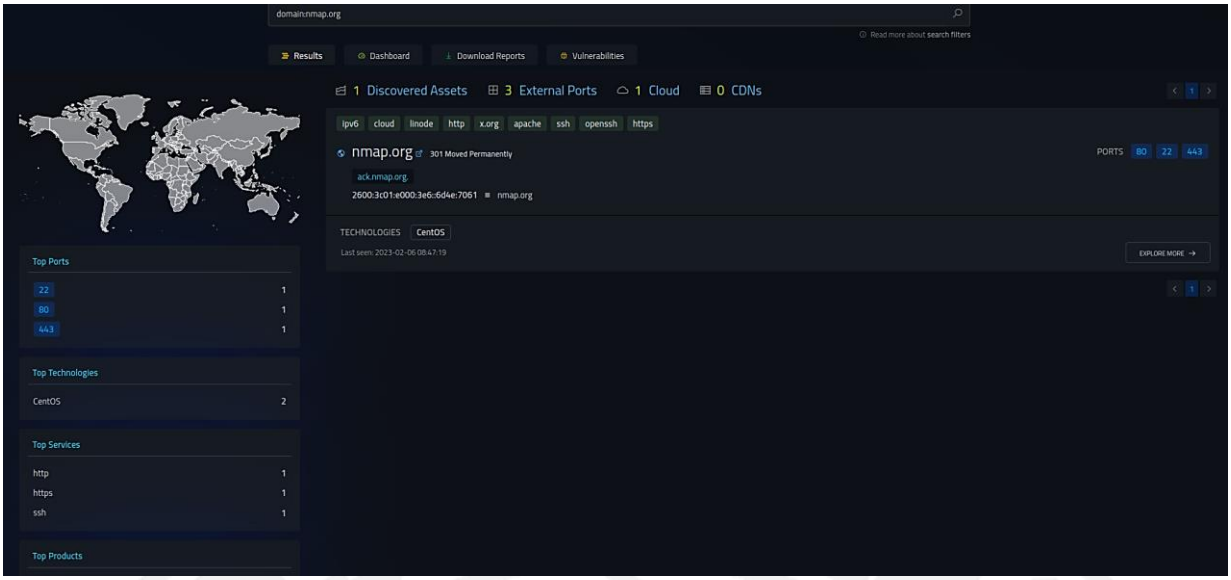


Figure 167. Example of a *FullHunt* search result for the domain *nmap.org*

GreyNoise (https://viz.greynoise.io/)

It searches for Internet-connected devices just like other engines in this section, working with IPs, CVE vulnerability identifiers (which devices have certain vulnerabilities), and others. Its main utility is to find out who has scanned any of your machines and know their intentions: if it is a targeted attack or simply one of the many web spiders that swarm the Internet of search engines and other applications of the style. You can find a list of possible information that you can find here: https://viz.greynoise.io/cheat-sheet, https://www.netmux.com/blog/greynoise-cheat-sheet

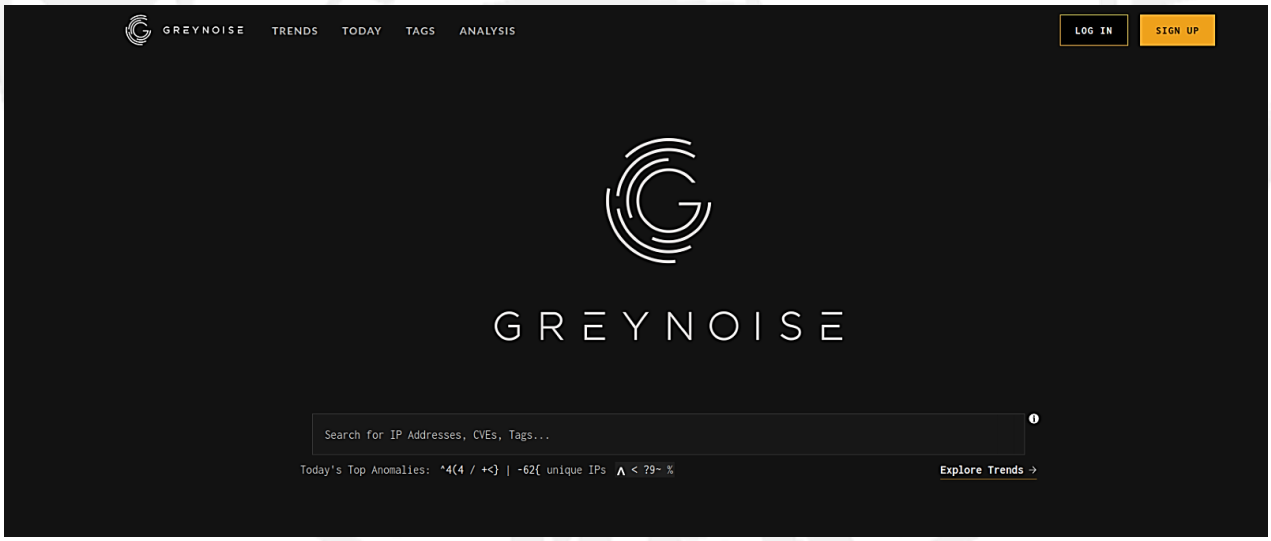


Figure 168. *GreyNoise*

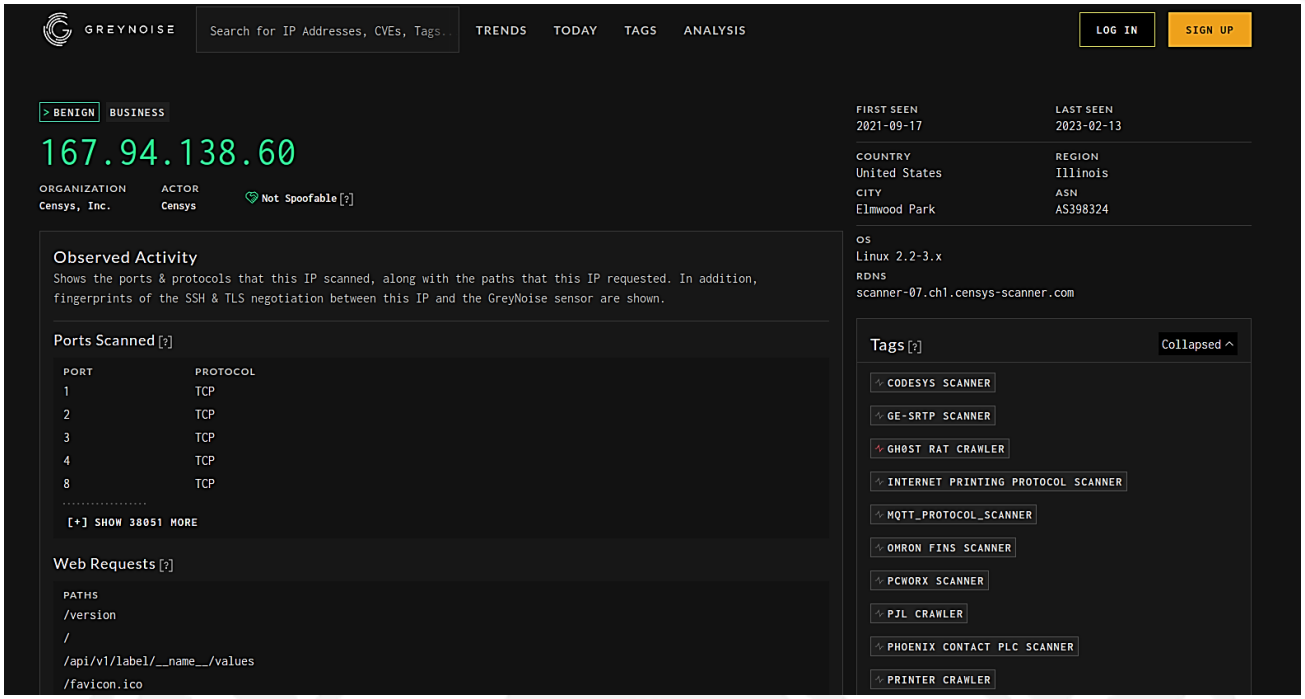


Figure 169. *GreyNoise* detecting a *Censys* scan by an IP that had in the log of a firewall

IVRE (https://ivre.rocks/)

Open-source framework for network recognition that is based on well-known tools of this type such as *Nmap*, *Masscan*, *ZGrab2*, *ZDNS* or *Zeek* to obtain data (network intelligence). It can be installed locally and saves the information in a configurable database, also incorporating tools to analyze them.

Figure 170. IVRE

Netlas (https://app.netlas.io/responses/)

It searches and monitors Internet-connected assets by scanning IPv4 addresses and crawling all known websites and web applications using protocols such as HTTP, FTP, SMTP, POP3, IMAP, SMB/CIFS, SSH, Telnet, SQL, and others. In https://twitter.com/Netlas_io/status/1719625320565813394 you can see many of their **common filters**:

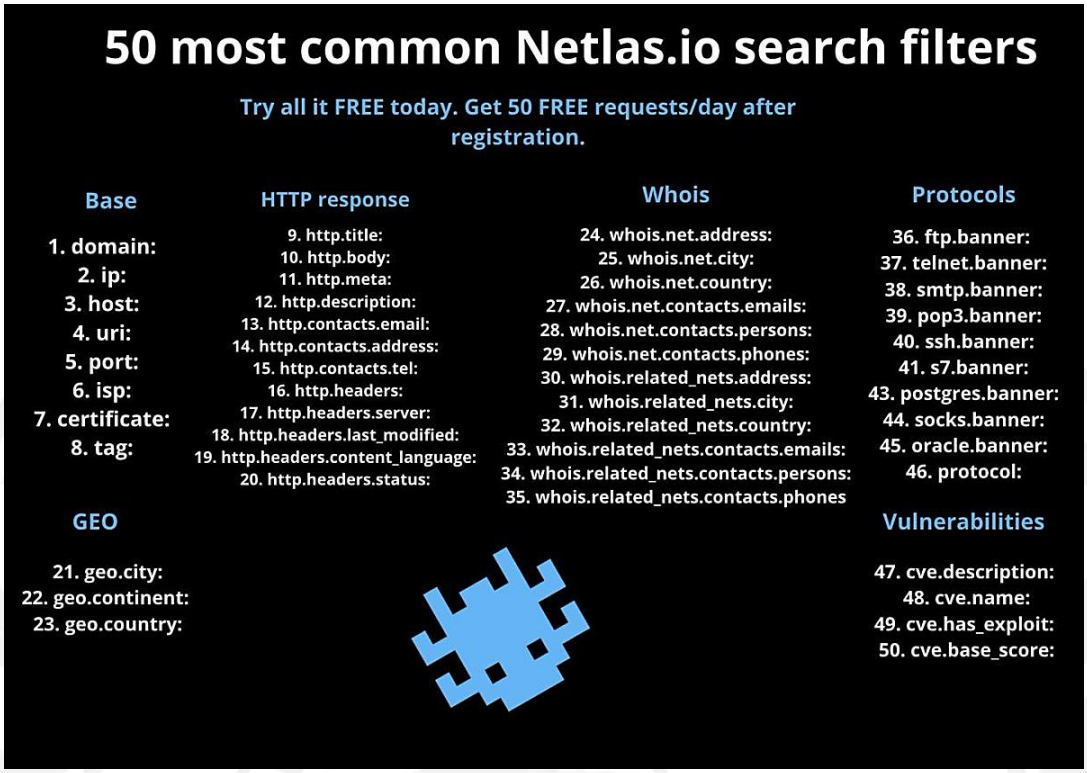


Figure 171. Netlas common filters

In addition, at this address we have a **complete reference of your queries**: https://github.com/netlas-io/netlas-cookbook

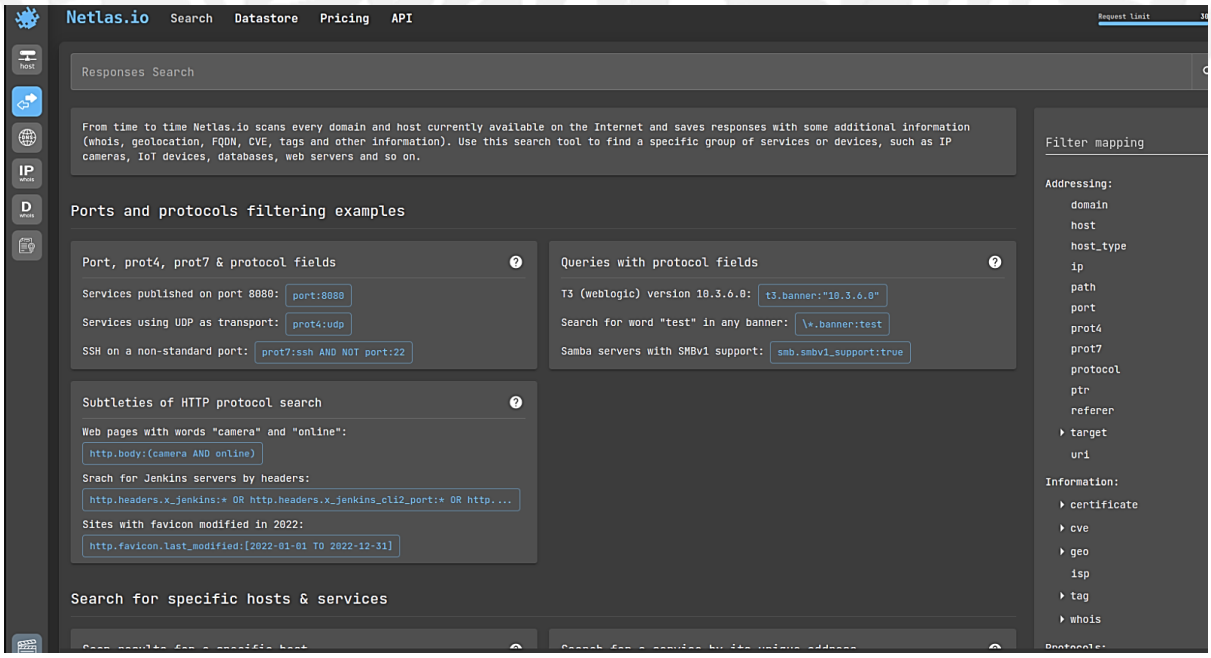


Figure 172. Netlas

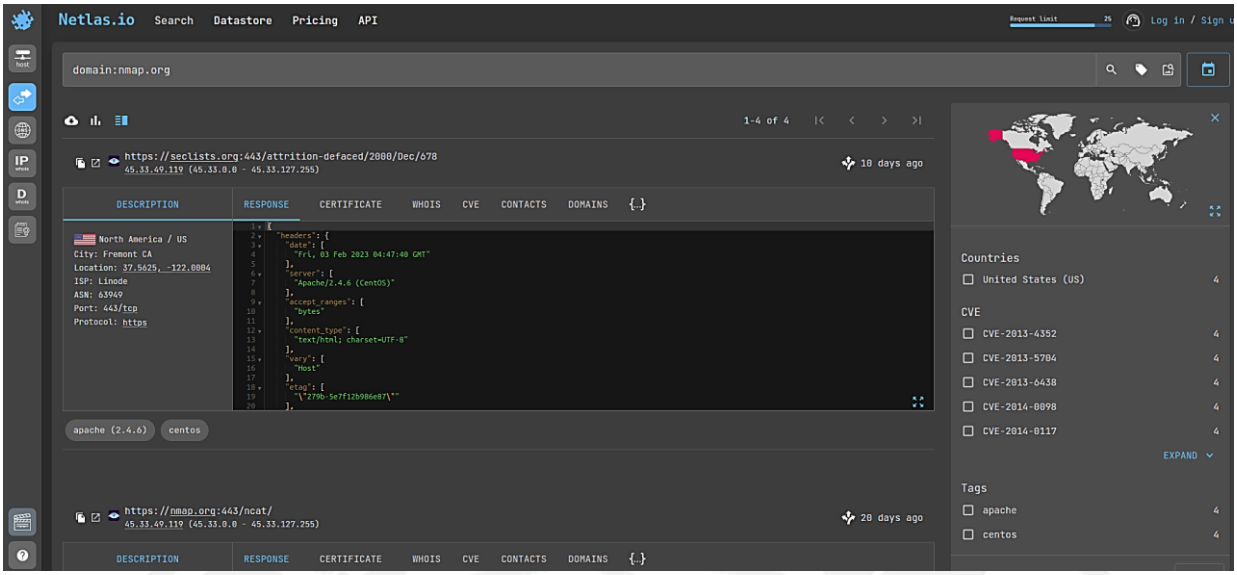


Figure 173. Example of using *Netlas* against the *nmap.org* domain

OnHype (https://www.onyphe.io/)

Cyber defense-oriented search engine for varied threat intelligence data collected by crawling from various sources available on the internet. In this link you can see a series of **typical queries** to get the most out of the engine: https://www.onyphe.io/files/onyphecheatsheetfreeview.pdf

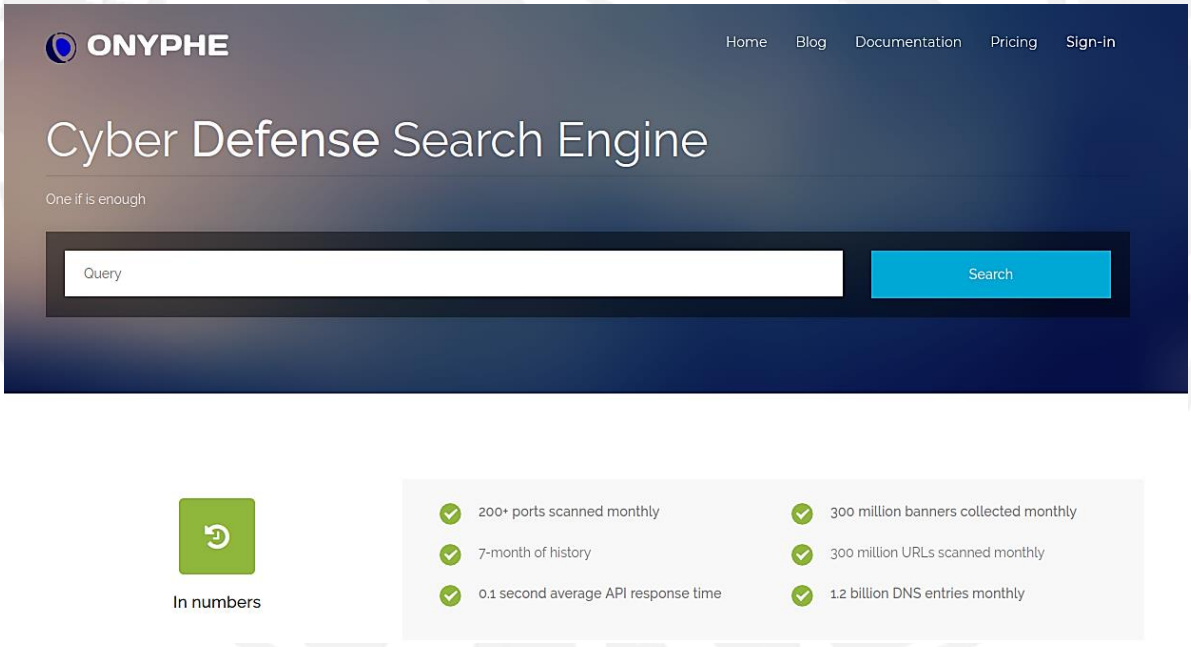


Figure 174. *OnHype*

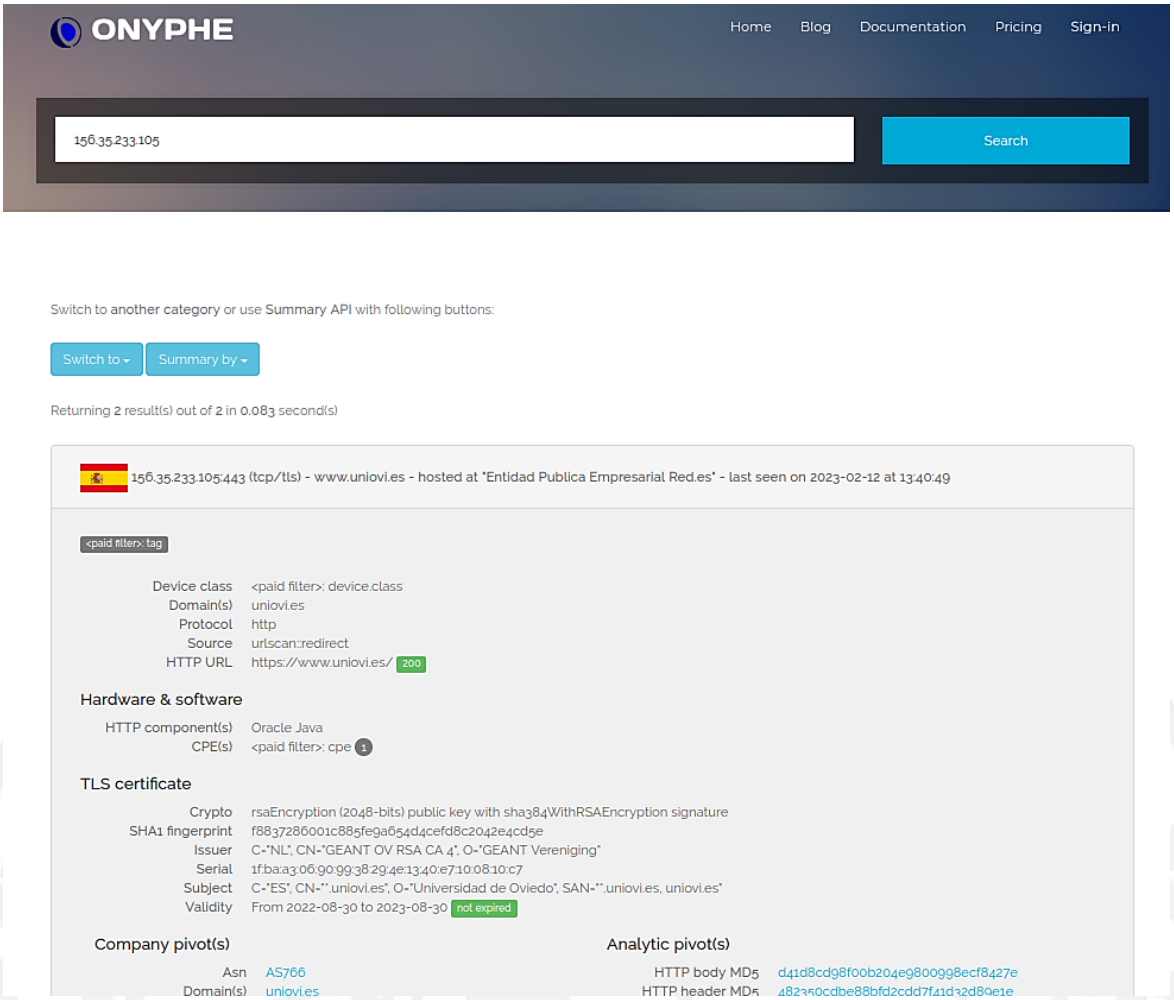


Figure 175. Example of *OnHype* search against uniovi.es

PulseDive (https://pulsedive.com/)

It offers quality threat intelligence by searching and investigating millions of IPs, domains, URLs, and threats, all for free, indicating whether or not they pose a security risk among other things.

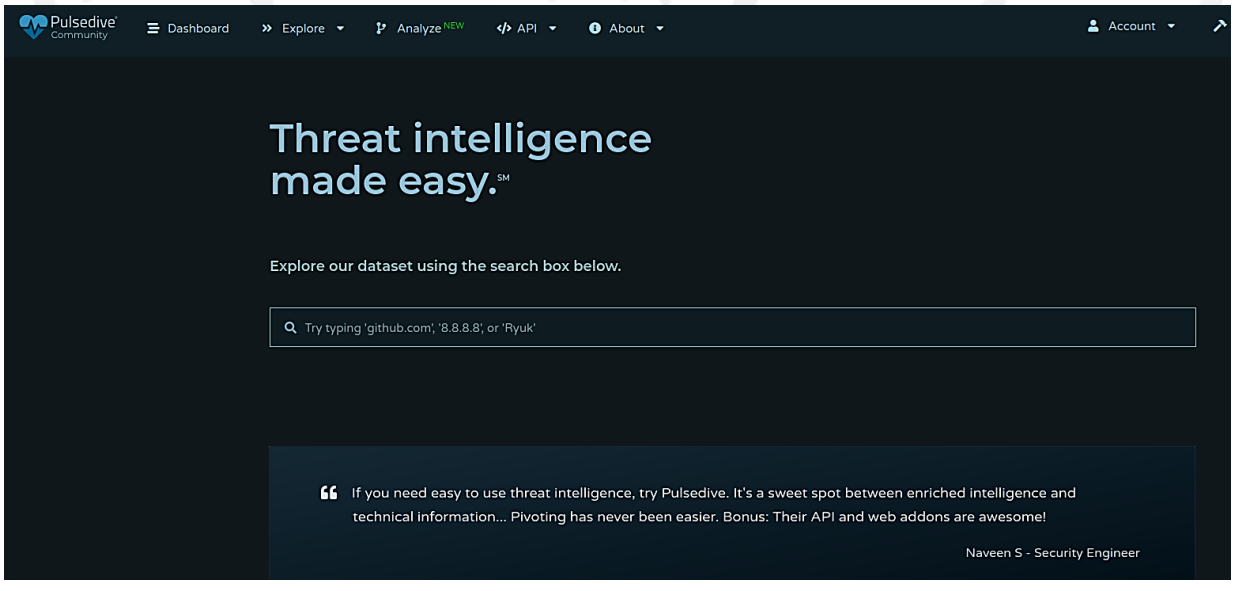


Figure 176. PulseDive

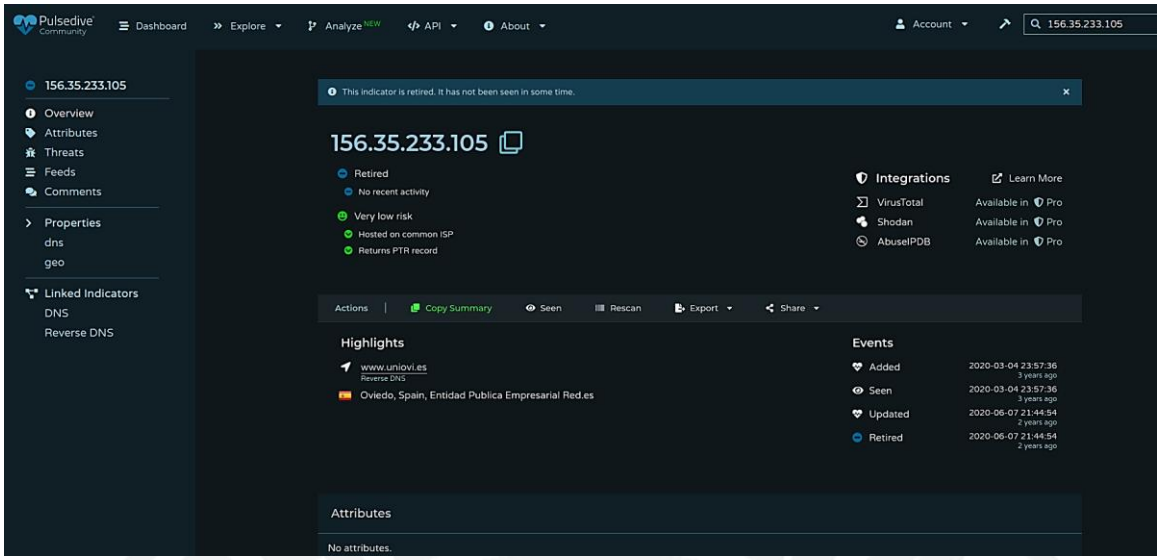


Figure 177. Example of using *PulseDive* against uniovi.es

Shodan (https://www.shodan.io/)

The popular research engine for conducting security research on internet-connected IoT devices that we saw in **Lab 2** and **Seminar 1**. This tool is used by thousands of security experts, researchers, CERTs, large organizations, and others around the world.

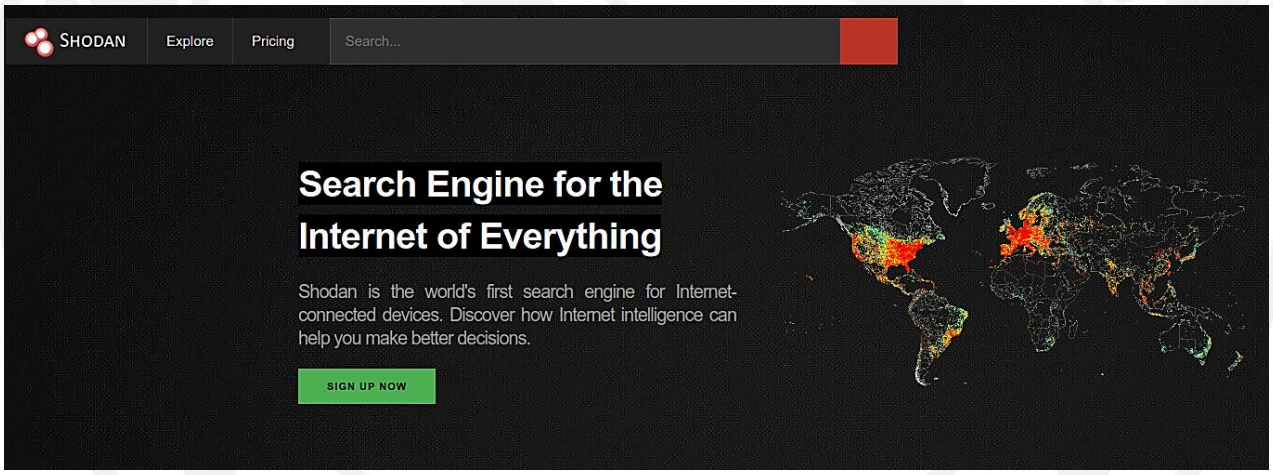


Figure 178. Shodan

The following *cheatsheet* summarizes the **most frequently asked queries** you can use on *Shodan*:

Cheatography

Shodan Cheat Sheet

by sir_slammington via cheatography.com/147122/cs/31955/

Common General Search Filters

ip:	Filter results by specific IP address.
asn:	Filter results by specific ASN ID.
hostname:	Filter results by specific hostname.
port:	Filter results by specific port number of service.
net:	Filter results from specified CIDR block.
isp:	Filter results by devices assigned a particular address (space) from a specified ISP.
city:	Filter results by specific city.
country:	Filter results by specific two-digit country code.
os:	Filter results by particular OS.
product:	Filter results by particular software.
version:	Filter results by specified version of software.

Common Premium API Search Filters

vuln:	Filter results by particular vulnerability ID (commonly CVE).
tag:	Filter results by tags on device.

HTTP Filters

http.component:	Filter results by a particular web technology.
http.status:	Filter results by specific status code.
http.html:	Filter results by strings found in HTML of files served.
http.title:	Filter results by string found in title of web pages served.

Common CLI Commands

count	Returns the number of results for a search.
domain	View all available information for a domain.
download	Download search results and save them in a compressed JSON file.
honeyscore	Check whether the IP is a honeypot or not.
host	View all available information for an IP address.
parse	Extract information out of compressed JSON files.
scan	Scan an IP/ netblock using Shodan.
search	Search the Shodan database.

Common CLI Search Fields

ip_str
port
org
hostnames
os
country
city

These will display their values upon a search, but won't provide statistics.

Common CLI Stats Facets

asn
city
country
cloud.provider
cloud.service
device
domain
ip
org
os

Common CLI Stats Facets (cont)

version

These will return statistical information about a given series of devices found on the public facing Internet. For example, it could be used to return the most common version found among devices running MariaDB in a particular ASN.

Use Case Examples

host: 8.8.8.8	shodan host 8.8.8.8	Display information about a Google's public DNS.
asn:15169 product:mysql	shodan stats asn:15169 product:mysql	Show information about devices within Google's ASN that run MySQL.
microsoft iis 6.0	shodan search --fields ip_str,port,org,hostnames microsoft iis 6.0	Detect IIS servers running on 6.0.
Navigate to https://honeyscore.shodan.io/ and enter target IP.	shodan honeyscore [TARGET]	Detect if given target is a honeypot or not.

Column one is the search you would perform in the Web UI. Column two is the search you would perform using the CLI utility, and the third column is an explanation of the search.

By sir_slammington
cheatography.com/sir-slammington/

Published 2nd May, 2022.
Last updated 2nd May, 2022.
Page 1 of 1.

Sponsored by Readable.com
Measure your website readability!
https://readable.com

Figure 179. Most frequent queries you can use on Shodan

Another one that is easier to use is this:

SHODAN Cheat Sheet

Cyber Writes

what is Shodan?

Shodan is a publicly available search engine which scans the entire Internet for a limited number of services and enumerates any discovered services by their banner responses, indexes that data and makes it searchable.

Shodan stores the information and indexes across five main fields: data, ip_str, port, org and location.country_code.

Be sure to use the 'View Raw Data' option on any discovered host to see all of the data Shodan has stored and learn possible new techniques of use.

While not always required, surround each search team in quotes to reduce confusion and broken queries.

IP Addresses & Subnets

Single IP Address - Search findings on single IP
Example: **52.179.197.205**

Hostname - Search for string in any hostnames
Example:
hostname:"microsoft.com"

Subnet - Search across a specific subnet range
Example: **net: "52.179.197.0/24"**

Port- Find any instances of active services on a port
Example: **port:"21"**

Service - Search for instances of specific services
Example: **"ftp"**

Service on Specific Port
Example: **"ftp" port:"21"**

Internet Service Provider - Search by ISP name
Example: **isp:"Spectrum"**

Autonomous System Number (ASN) - Search by ASN
Example: **ASN:"AS8075"**

Physical Location

Country - Search by country code
Example:
country :"US"

City - Search by city name
Example:
city :"New York"

State - Search by state code abbreviation
Example:
state :"NY" or region :"NY"

Zip Code - Search by postal ZIP code
Example:
postal:"92127"

Geo - Search by GPS coordinates
Example:
geo:"40.759487,-73.978356"

Geo - Search by GPS (within a range of 2 km)
Example:
geo:"40.759487,-73.978356,2"

Operating Systems, Products

Operating System - Search by operating system type
Examples: **os:"Windows Server 2008"**
os:"Linux 2.6.x"

Organization/Company - Search by organization name
Example: **org:"Microsoft"**

Product - Search by known product name
Example: **product:"Cisco C3550 Router"**

Version - Search for specific version number
Example: **product:"nginx" version:"1.8.1"**

Category - Search by Shodan category
Example: **category:"ics" or category:"malware"**

Microsoft SMB - Search for specific SMB versions
Example: **smb:"1" or smb:"2"**

Microsoft Shared Folders - Find exposed shared folders
Example: **port:"445" "shares"**

Web Apps

Page's Title - Search for text in page's title
Example: **title: "Index of /ftp"**
Page's HTML Body - Search body of webpage for text string
Example: **html:"XML-RPC server accepts"**
Web Technologies - Search for specific web technologies
Example: **http.component:"php"**
SSL/TLS -Search for SSL/TLS versions supported
Example: **ssl.version:"sslv3" or ssl.version:"tlsv1.1"**
Expired Certificates - Search for expired HTTPS certs
Example: **ssl.cert.expired:"true"**

Other

Date: After - Search for findings that appear after a date
Example: **after:"01/01/18"**

Date: Before - Search for findings that appear before a date
Example: **before:"12/31/17"**

Screenshot - Display results which only have screenshots
Example: **port:"80"**
has_screenshot:"true"
*** Watch the webcams roll in!**

port:"3389"
has_screenshot:"true"
***Watch for exposed Window domain & users!**

Limited Access

There are number of useful operators that require premium paid accounts (Enterprise, Academic, etc)

Vulnerability - Search by CVE ID number
Example: **vuln:"CVE-2017-0143"**

Tags - Search based on Shodan tagged data
Example:
tag:"ics" or tag:"database"

Figure 180. More frequent queries you can use on *Shodan*

Wigle (https://wigle.net/)

WiGLE (or *Wireless Geographic Logging Engine*) is a website for **collecting information about the different wireless access points around the world**. Users can register on the website and upload access point data such as GPS coordinates, SSID, MAC address, and the type of encryption used on the discovered access points. It is like Warchalking (**Seminar 2**), but on steroids 😊

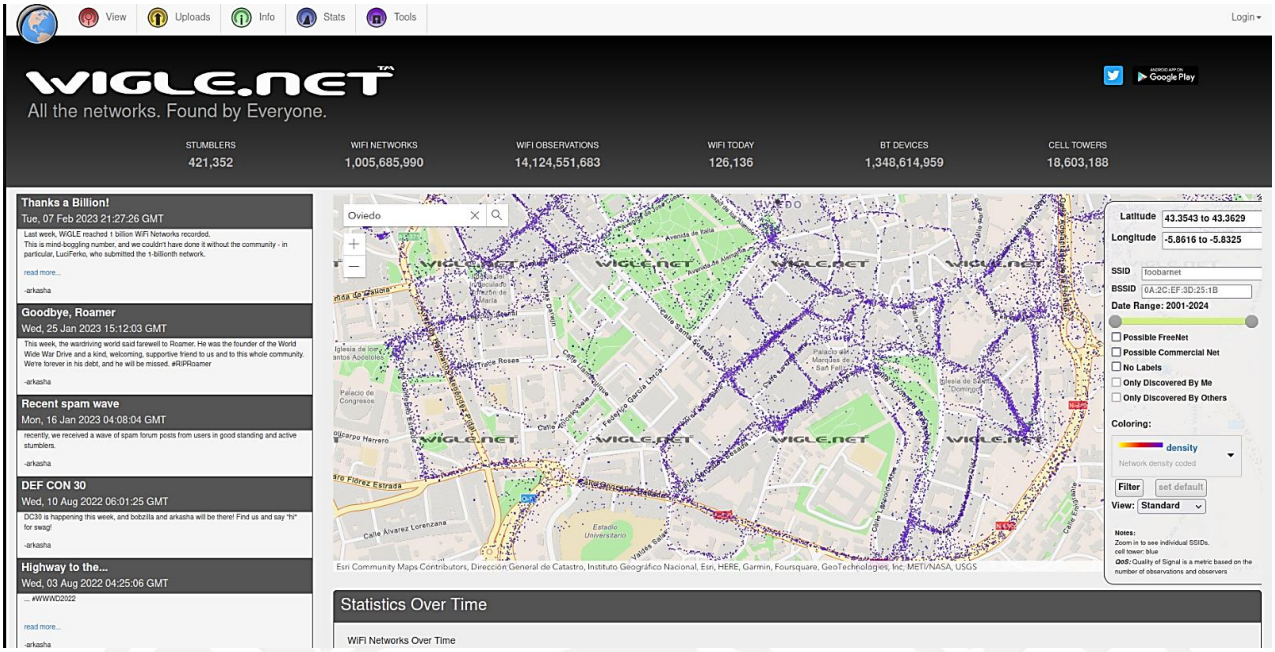


Figure 181. *Wigle* (showing Wi-Fis from Oviedo)

ZoomEye (https://www.zoomeye.org/)

It maps machines and their topography locally or globally based on IPv6, IPv4 and domain name databases. Find active service protocols and ports at a given instant of time using a simple query language. It works for example on IPs and files and **requires free registration**. This file contains a number of **typical useful queries**: https://github.com/interference-security/zoomeye-data/blob/master/usage.txt

Figure 182. ZoomEye

Search engines for websites or networks

Advance Online Utilities (https://centralops.net/co/)

In this address we can find various website exploration tools

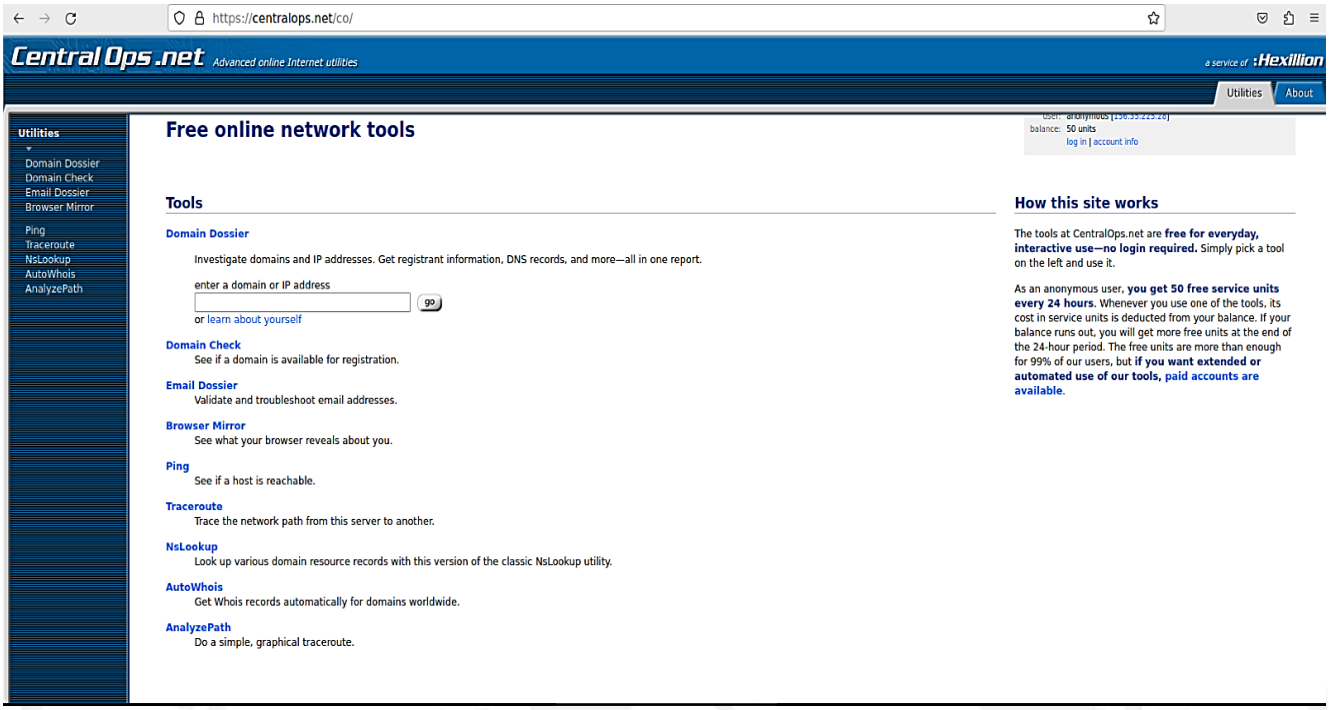


Figure 183. ***Advance Online Utilities***

Cisco Reputation Check (https://www.talosintelligence.com/)

Web that checks the reputation of a URL or IP we provide

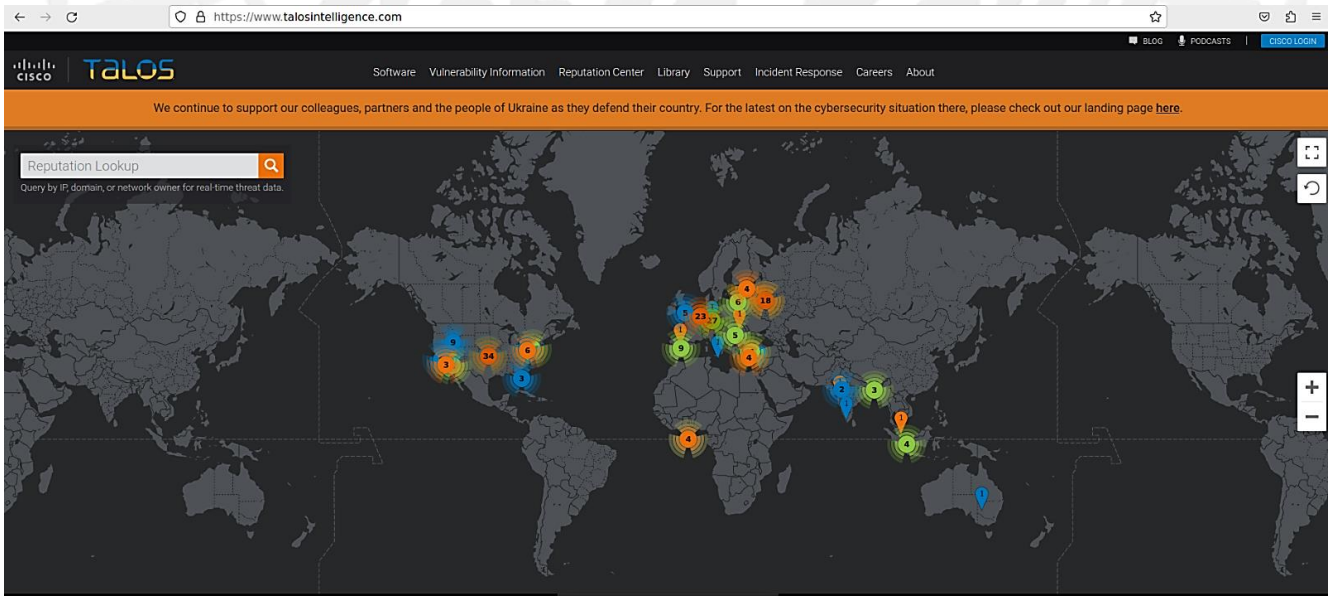


Figure 184. ***CISCO Reputation check***

Cloudflare Radar URL Scanner (https://radar.cloudflare.com/scan)

Search engine that from a URL makes you a report with many technical details of the web that is behind it: if it has indications of being phishing, data from its TLS certificate, data from requests and responses, DNS records, if it deals with cookies with security, technologies and libraries used ...

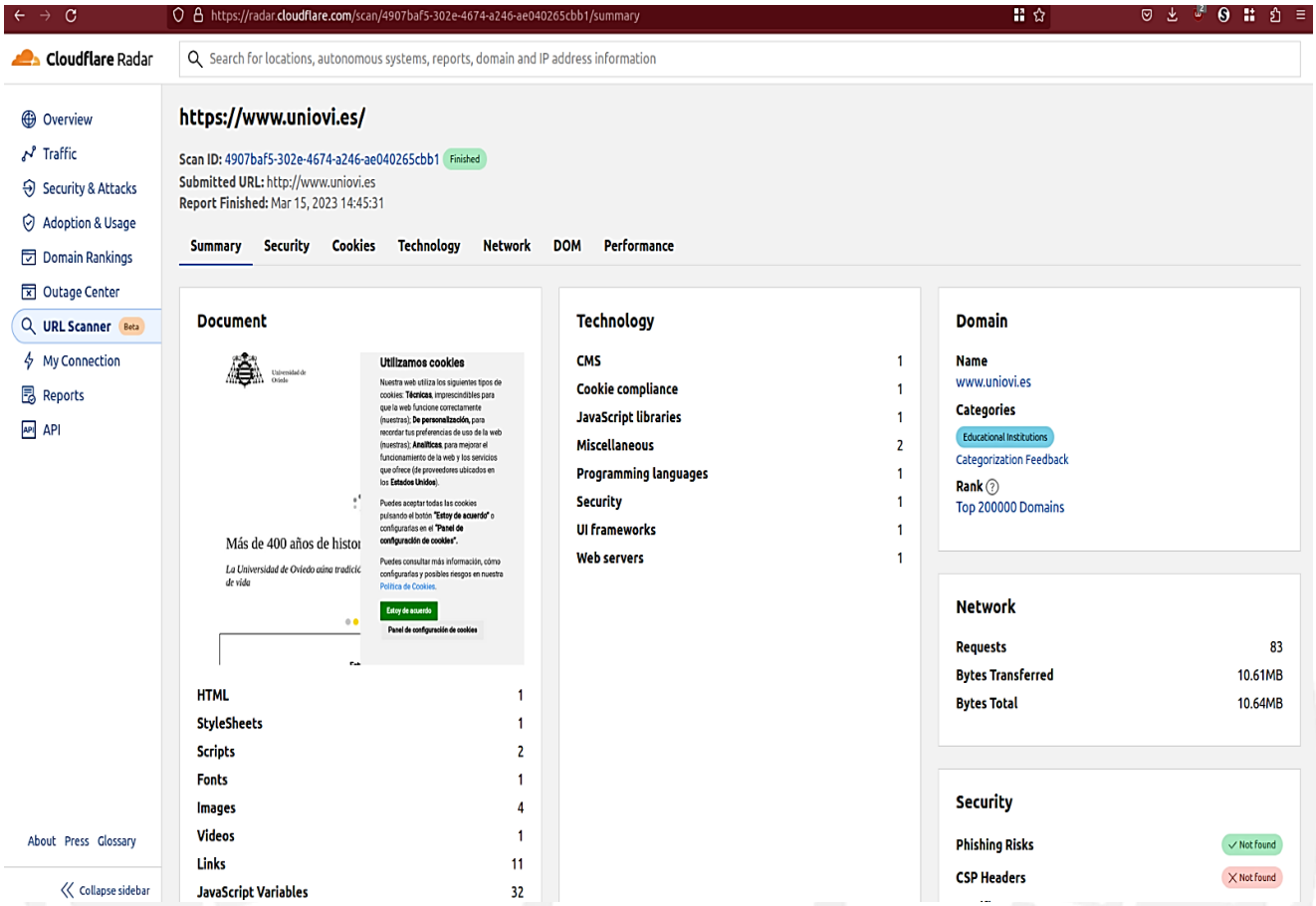


Figure 185. ***Cloudflare Radar URL Scanner***

CyberGordon (https://cybergordon.com/)

Web that checks the online reputation of an IP or web to know if it can be dangerous or not.

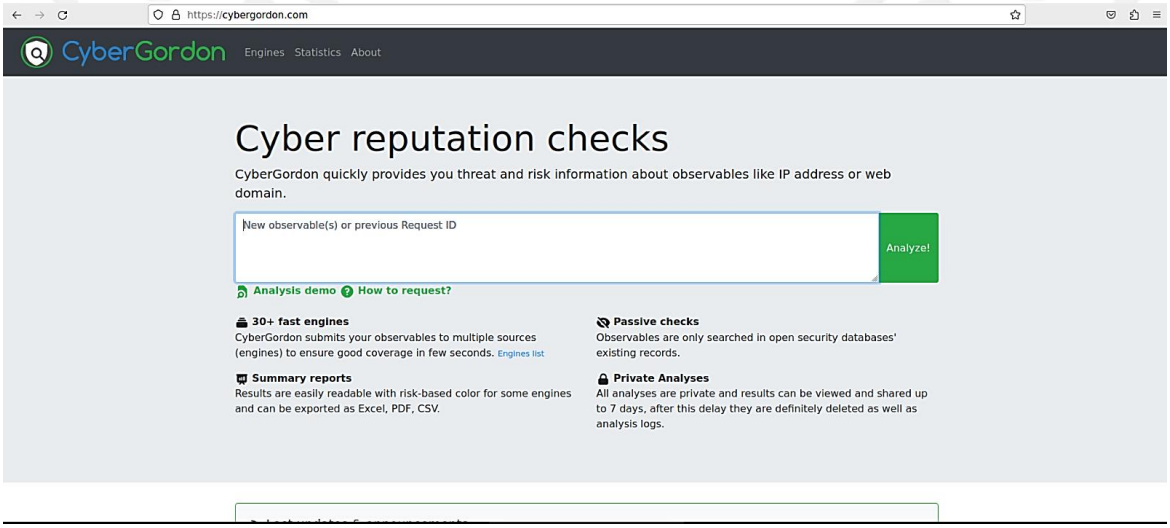


Figure 186. ***CyberGordon***

IBM Reputation Check (https://exchange.xforce.ibmcloud.com/)

Another website that checks the reputation of others. Although it is not free, it is very popular in business environments (in case you need a website like this in your future work)

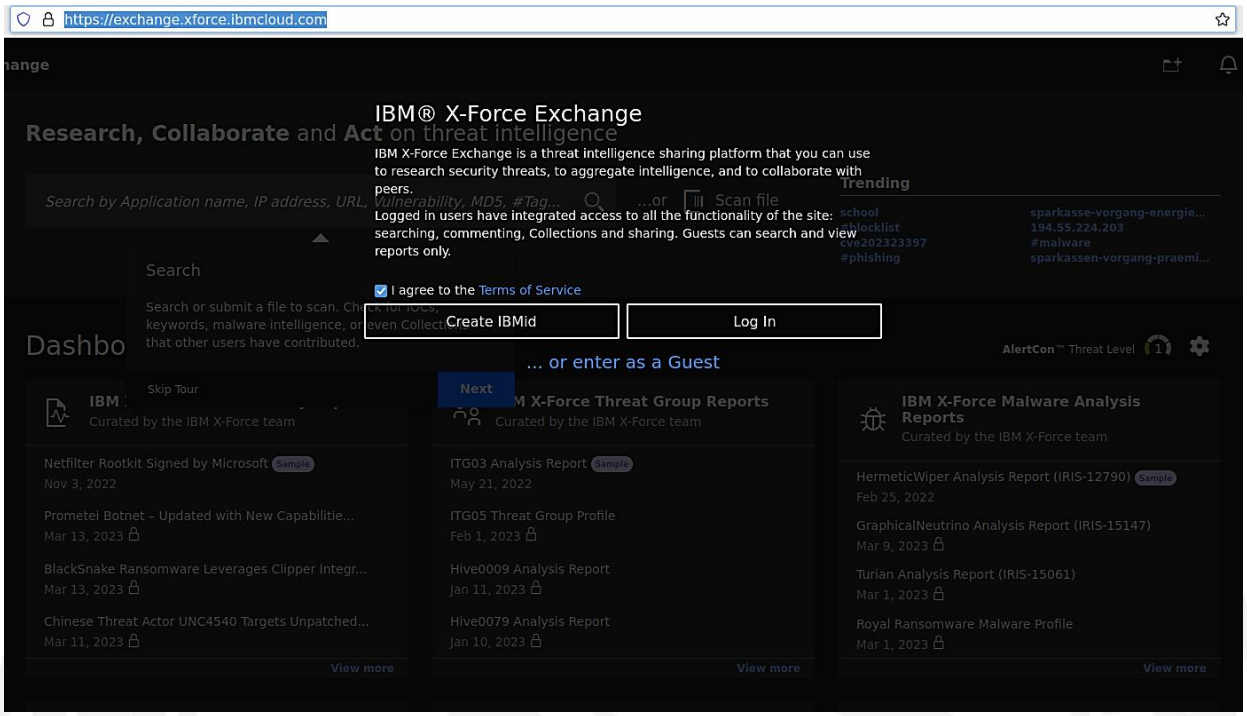


Figure 187. *IBM X-Force Exchange*

IP Reputation Check (https://www.abuseipdb.com/)

Another website that allows you to know if an IP has been identified as the source of spam, attacks, and other forms of online abuse.

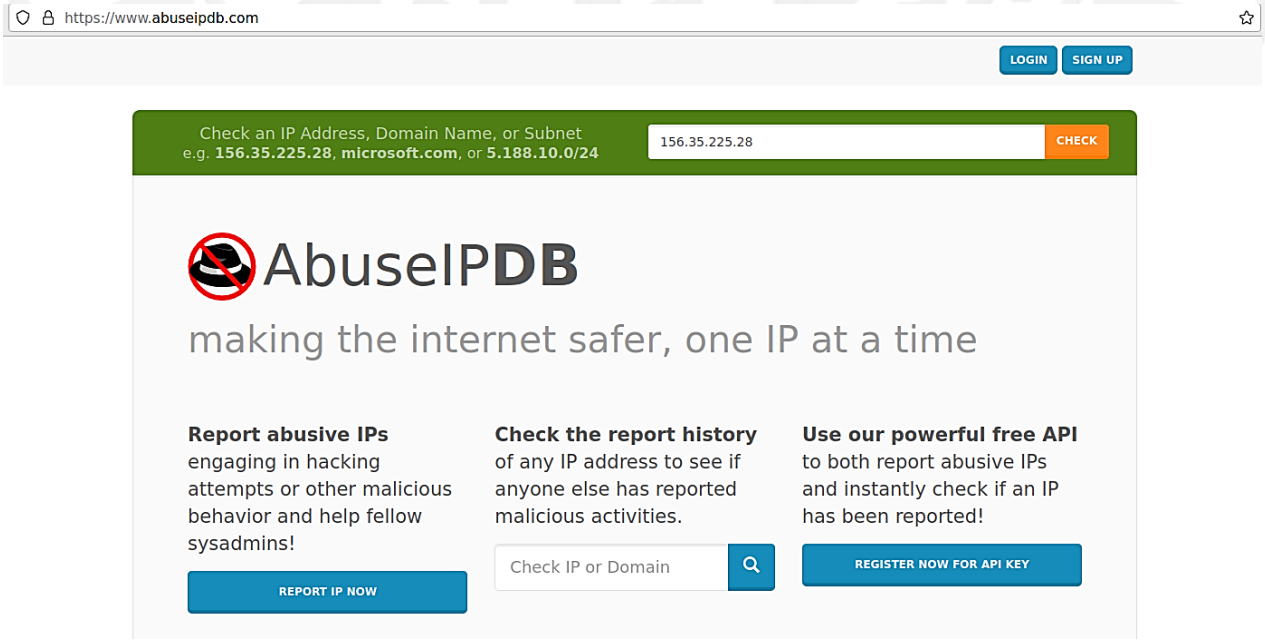


Figure 188. Abuse IP DB

PublicWWW (https://publicwww.com/)

It is a digital and affiliate marketing research tool that finds any alphanumeric snippet, signature or keyword in the HTML, JS, and CSS code of web pages. It is a useful tool for digital marketing and affiliate marketing research, but also for locating certain pieces of key code on different websites.

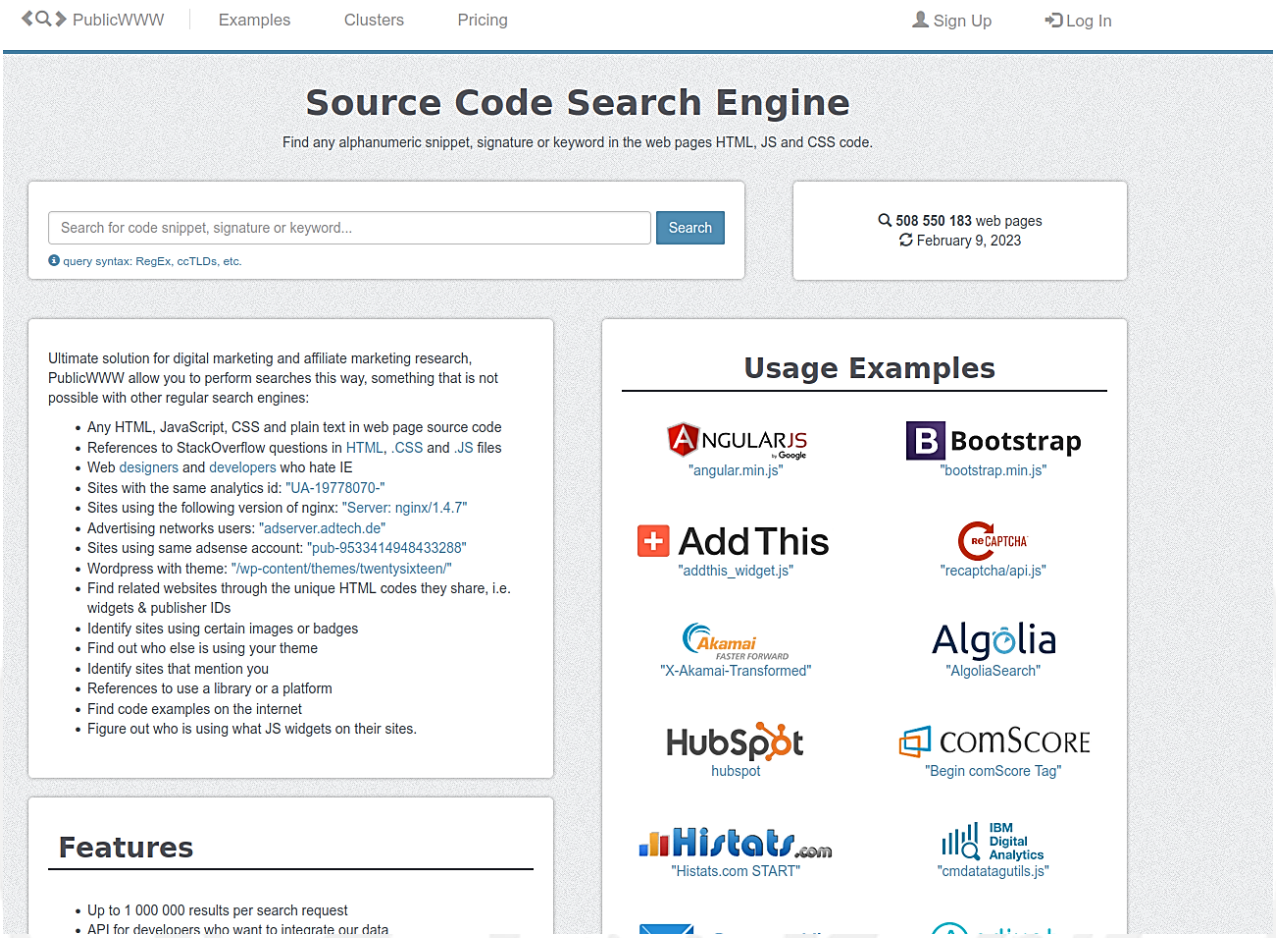


Figure 189. *PublicWWW*

URL Category Finder (https://sitereview.bluecoat.com/#/)

This website allows you to find out what type of website is the one that is served from a URL that we provide, to know what we can find if we visit it.

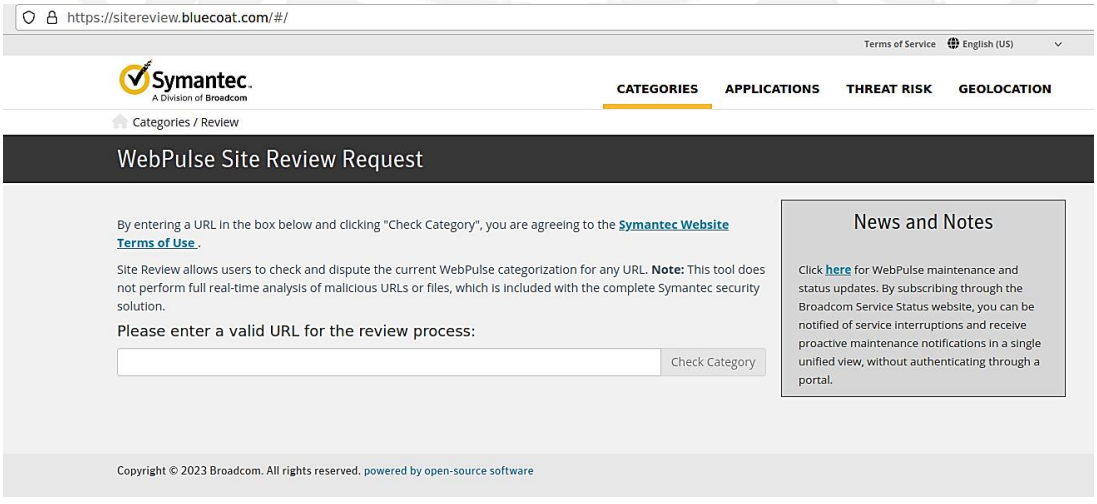


Figure 190. URL Category Finder

URLScan (https://urlscan.io)

Free service to scan and analyze websites. It has automated processes that will navigate to the URLs provided as a normal user and record the activity that this page navigation creates, offering varied information but in a way that our machine is isolated from navigation.

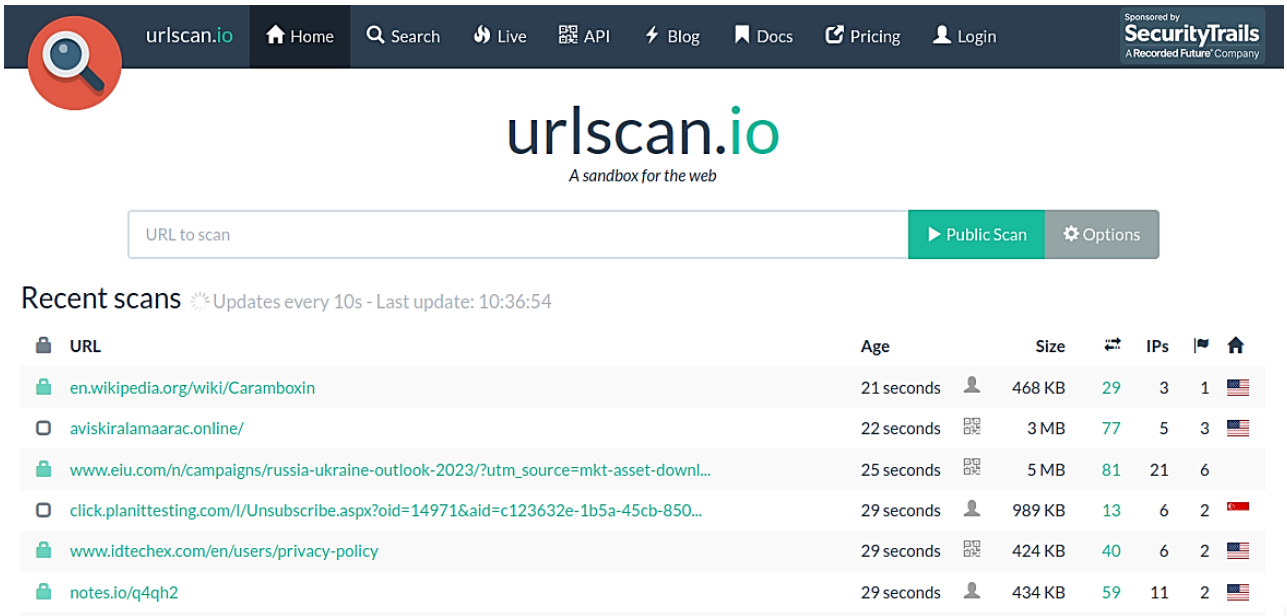


Figure 191. *URLScan*

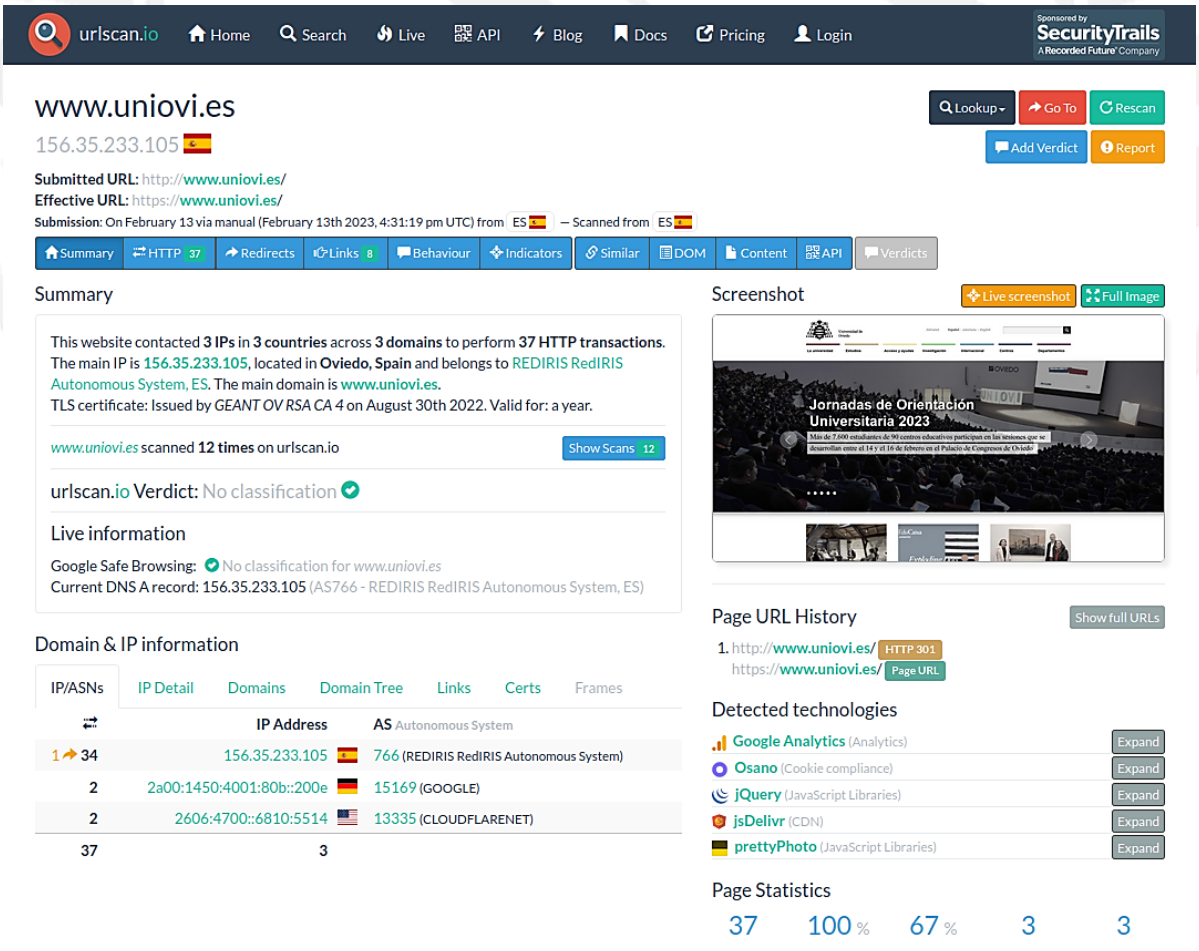


Figure 192. URLScan against uniovi.es

Website Profiler Tool (https://builtwith.com/)

This website allows us to identify the technologies that have been used to build the websites whose URL we provide, which can then be used to continue our investigation of the owner.

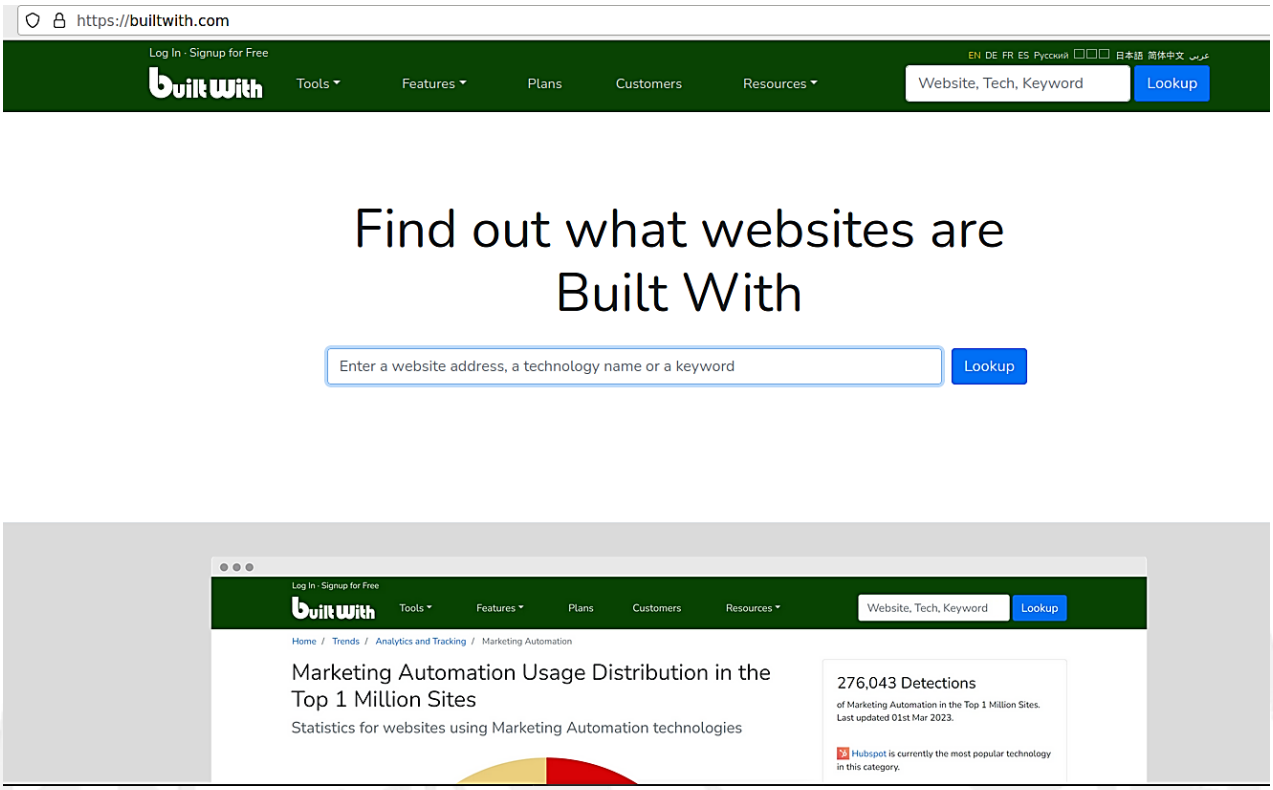


Figure 193. Website Profiler Tool

Vulnerability search engines

CyberChef (https://gchq.github.io/CyberChef/)

Itis a tool to manipulate data in many different ways: encoding, cryptography... **It is not in itself a search engine for vulnerabilities**, but it can be used to discover vulnerabilities from processing things found by other engines, by really revealing what they are and therefore being able to pose a danger to the web that has them.

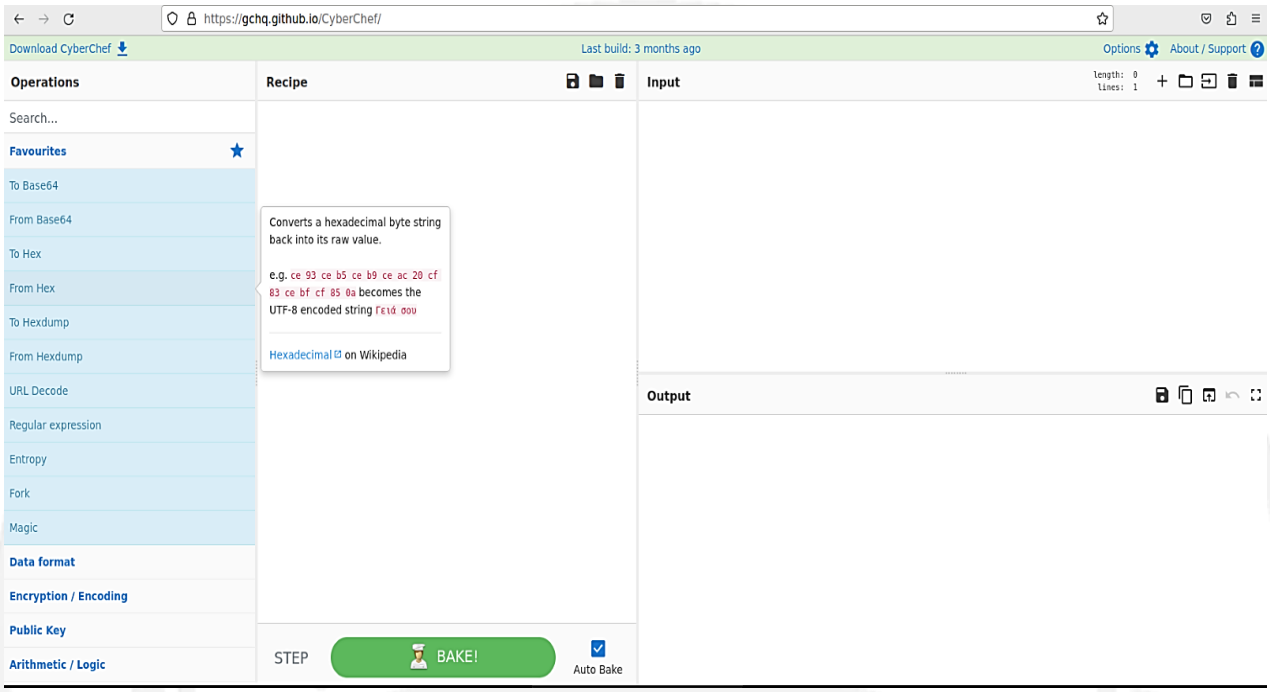


Figure 194. ***CyberChef***

Packet Storm Security (https://packetstormsecurity.com/)

Website that offers current and historical computer security tools, exploits and security advisories. Operated by a group of security enthusiasts who post new security information and offer tools for educational and testing purposes.

Figure 195. ***Packet Storm Security***

Vulners (https://vulners.com/)

Security database that contains descriptions of a large number of software vulnerabilities in a machine-readable format and that keeps up to date with the latest security threats, also allowing, after registration, to scan vulnerabilities to the IPs that we specify.

Figure 196. ***Vulners***

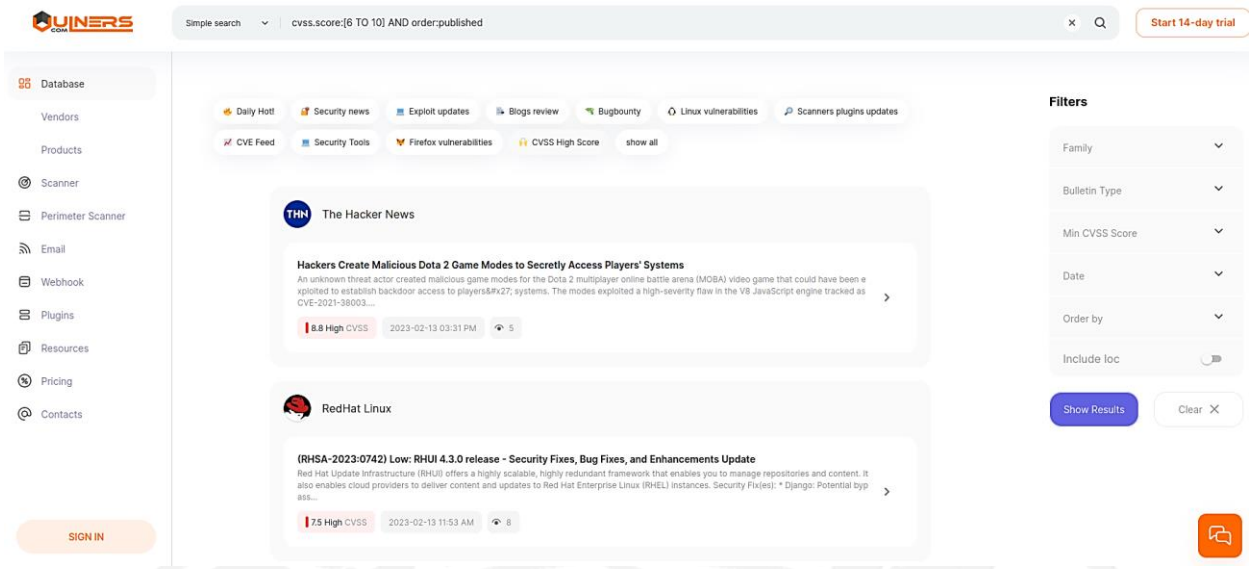


Figure 197. Known vulnerability search example with *Vulners*

Search engines for general data breaches

CRT.sh (https://ui.ctsearch.entrust.com/ui/ctsearchui)

Search engine for certificates that have been registered by Certificate Transparency (CT). It is a site where you can find all the SSL or TLS certificates of the particular target domain and thus **query the data that was used when registering them**, which may have names and other information of the target that can be used in other OSINT operations. It is also an effective way to discover target domains that may not be visible otherwise.

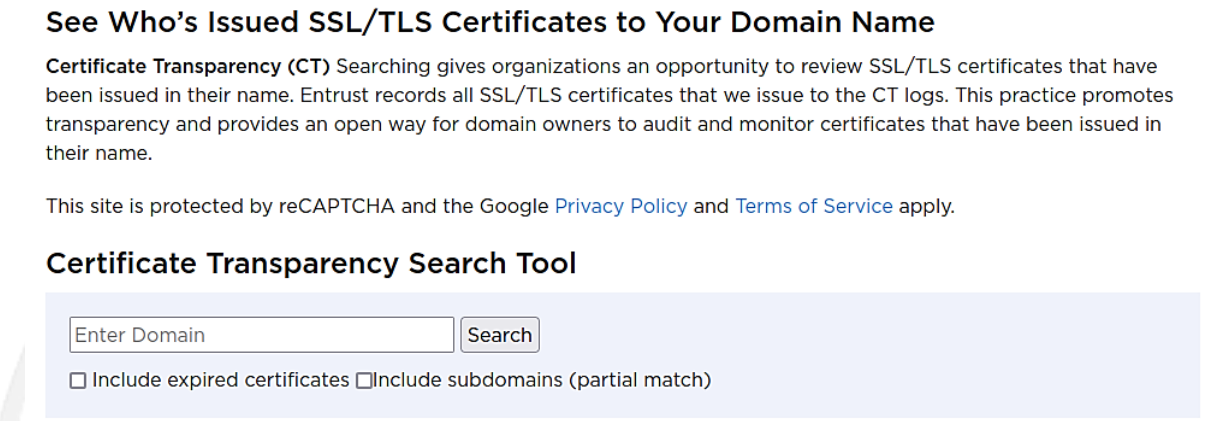


Figure 198. *CRT.sh*

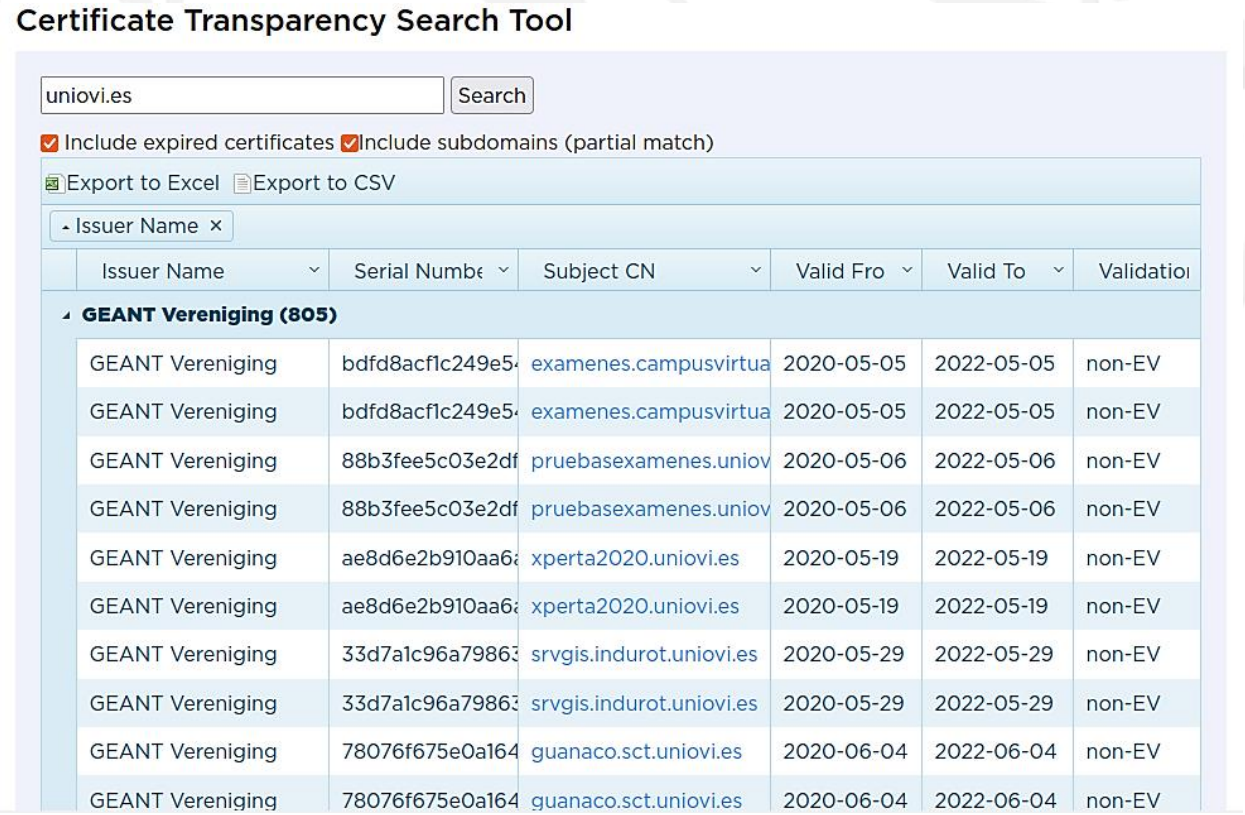


Figure 199. Search for certificates associated with uniovi.es with *CRT.sh*

FilePursuit (https://filepursuit.com/)

A search engine for all types of files on the Internet, which obviously can correspond to leaks of a target that we are studying

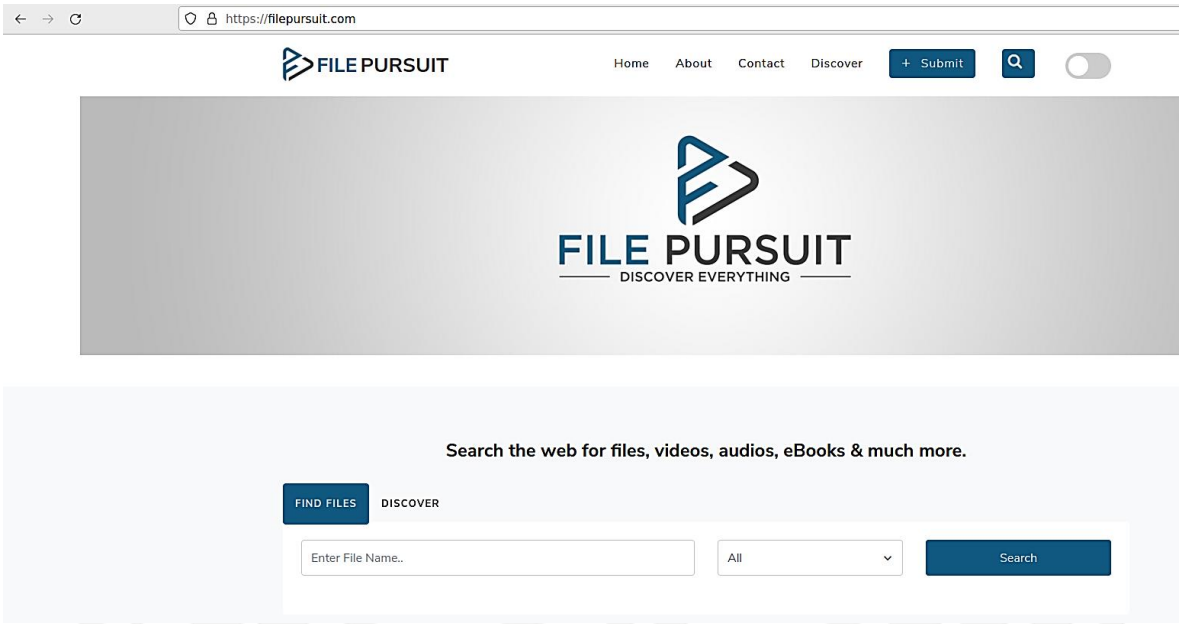


Figure 200. *FilePursuit*

Gray Hat Warfare (https://buckets.grayhatwarfare.com/)

This web tool allows you to easily find open *buckets* on cloud providers (e.g., Amazon S3) with potential owner information. It also allows you to find other documents quickly by searching for several file types.

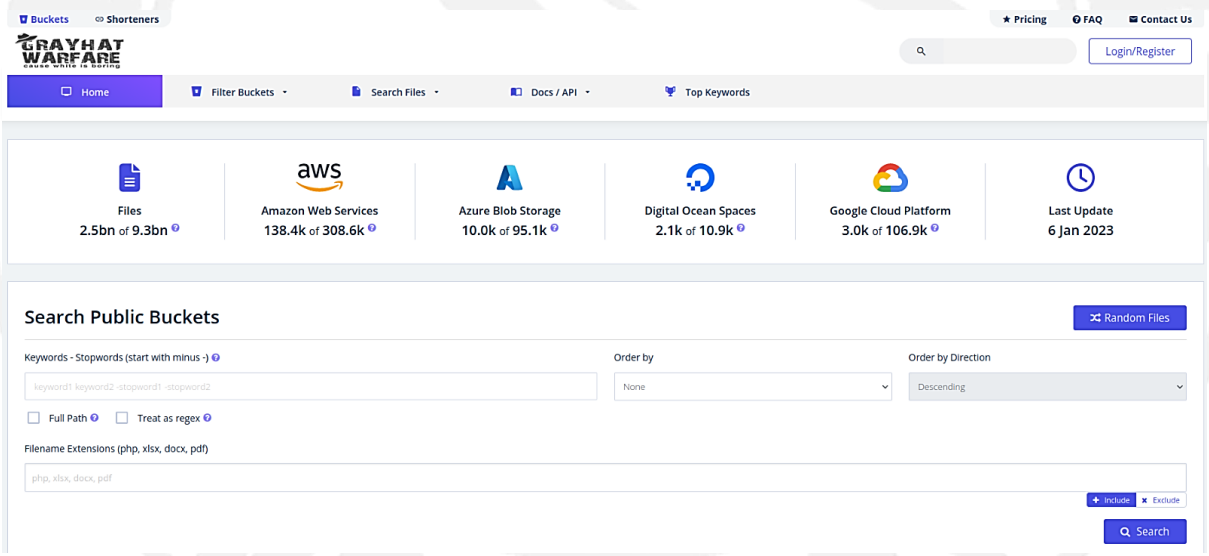


Figure 201. *Gray Hat Warfare*

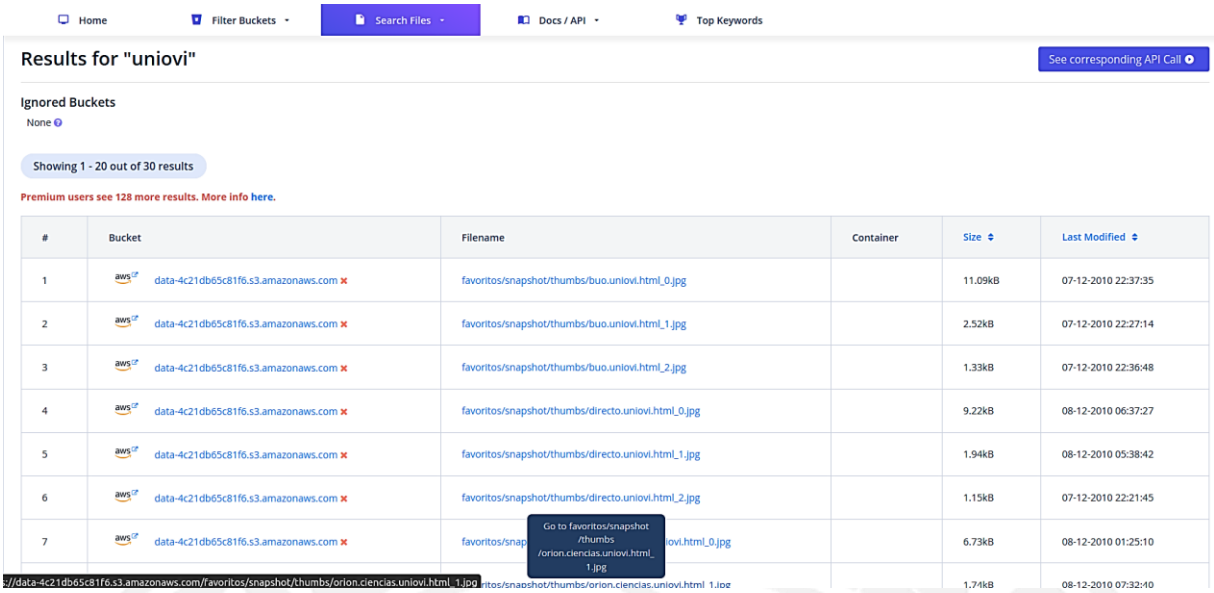


Figure 202. Search for buckets associated with Uniovi with *Gray Hat Warfare*

IntelligenceX (https://intelx.io/tools)

It is a collection of tools to search third-party sites and extract intelligence from social networks such as *Facebook*, *Twitter*, *LinkedIn*, *Telegram*, and *YouTube*. It also searches for *ToR*, *I2P*, data leaks, domains, and emails. In other words, it functions as an aggregator of many useful search services (some included in this document and in the course and others not, since there are many) so that their use can be coordinated.

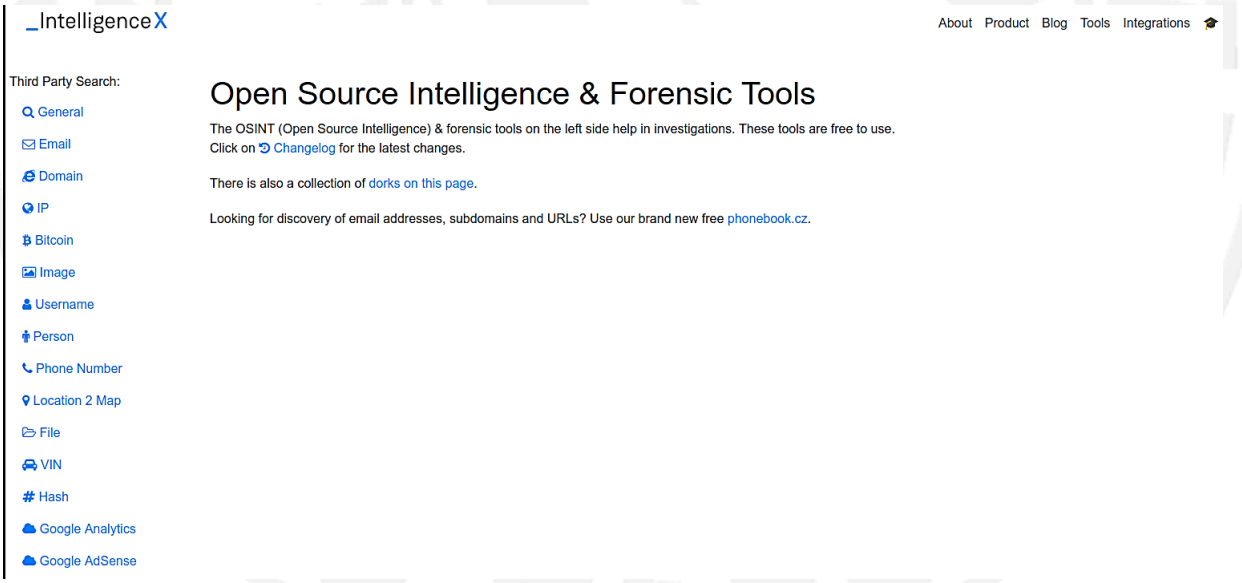


Figure 203. *IntelligenceX*

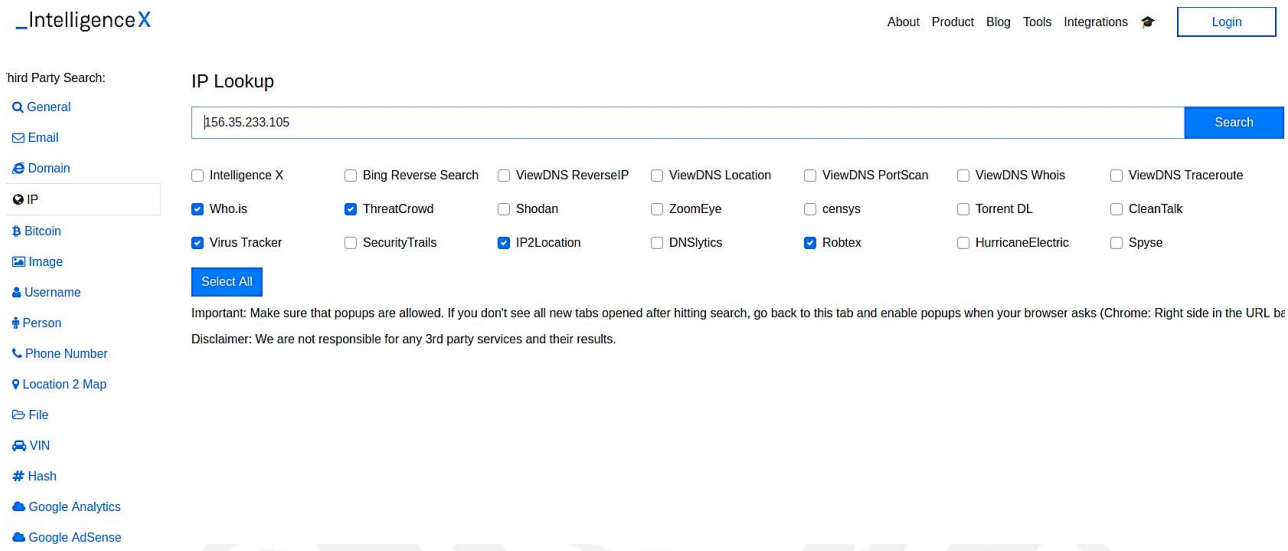


Figure 204. *IntelligenceX* and its service catalog on a given IP (Uniovi)

LeakIX (https://leakix.net/)

Platform that combines a search engine that indexes public information and an open reporting platform linked to the results obtained by these searches.

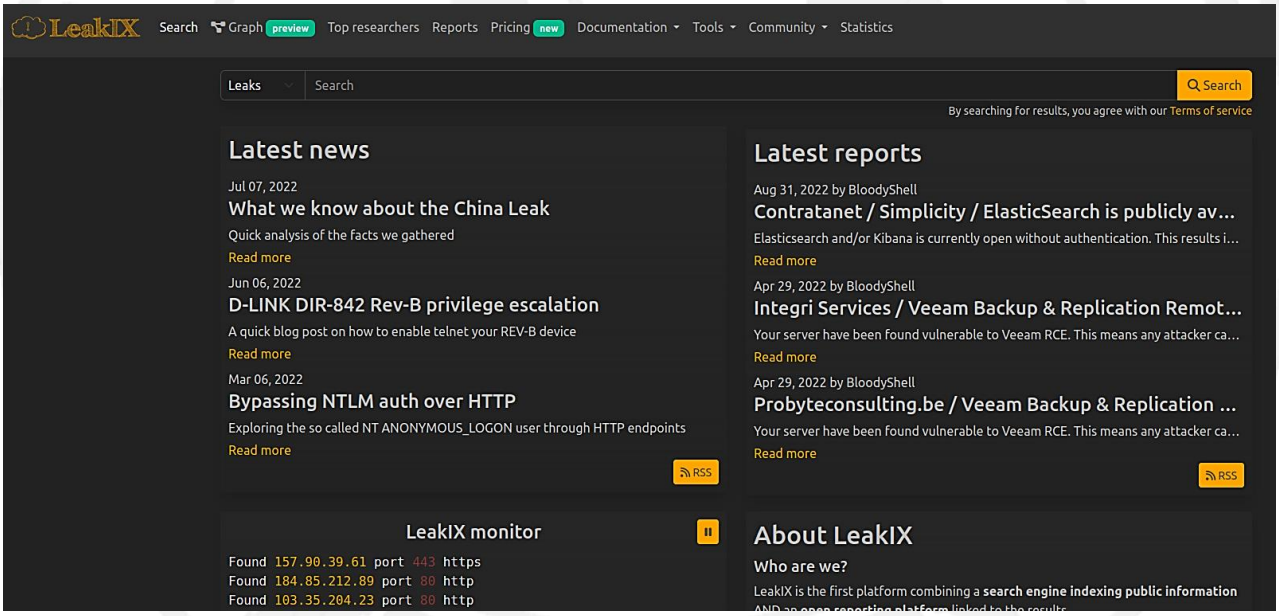


Figure 205. LeakIX

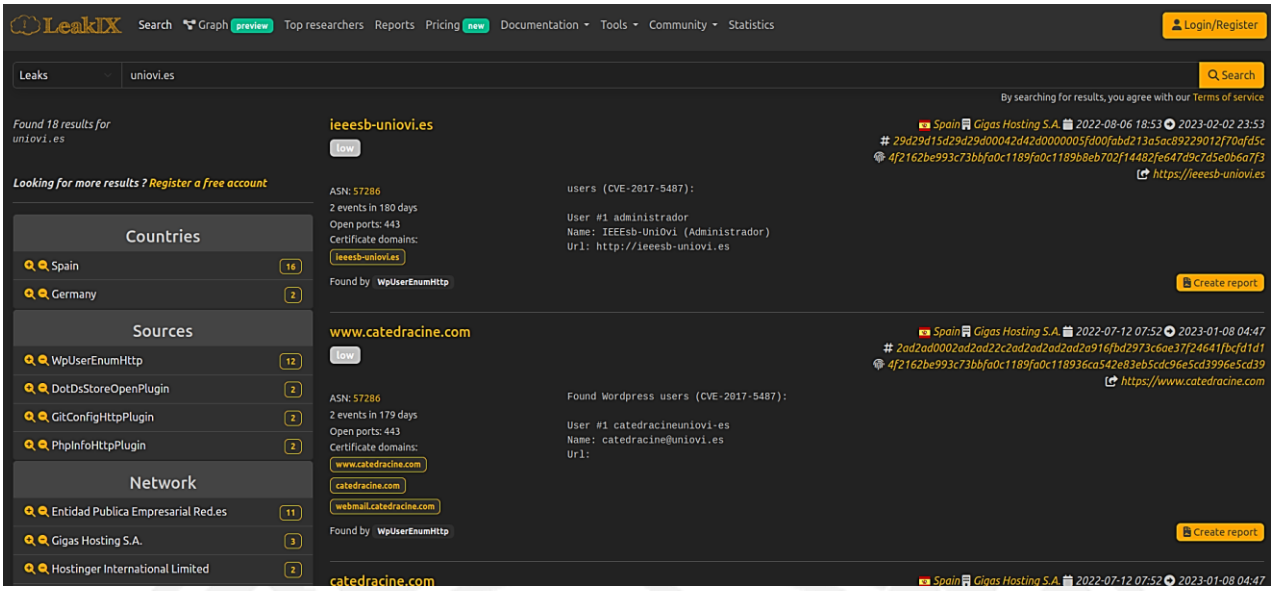


Figure 206. Data breaches associated with uniovi.es located by *LeakIX*

For each data leak, *LeakIX* also offers the reason and the data that has been leaked (in this case valid users on a website made with *WordPress*).

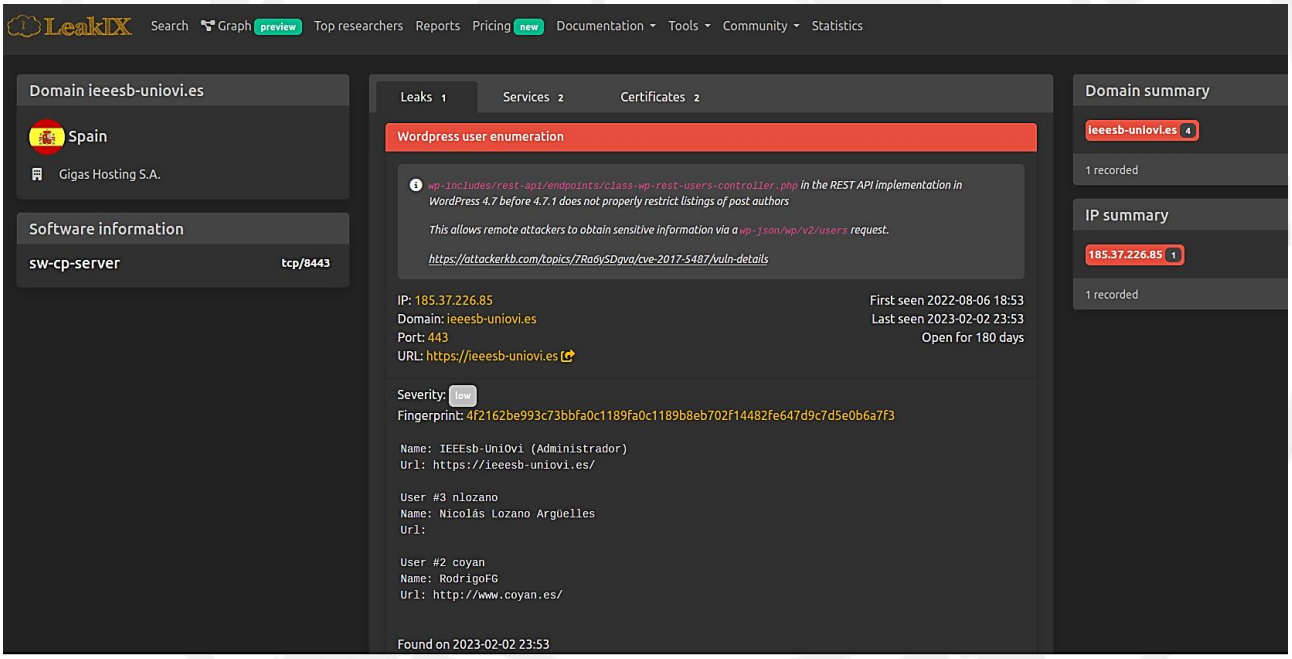


Figure 207. Description of a leak found with *LeakIX*

TinEye Reverse Image search (https://tineye.com)

An alternative service to Google's famous reverse image search, with the same image search functionality.

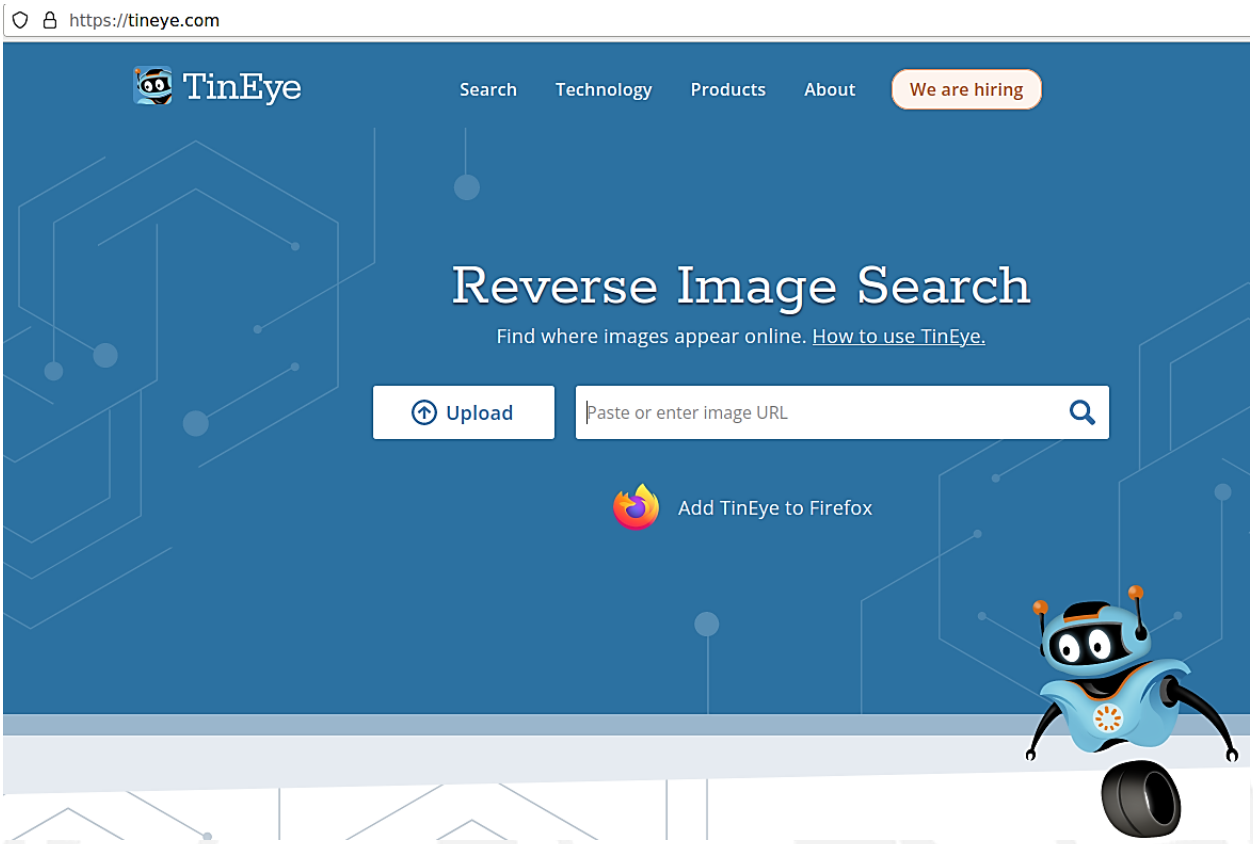


Figure 208. ***TinEye***

Specific search engines for leaks related to user accounts

Hunter.io (https://hunter.io/)

Mentioned in **Seminar 1**, look for email addresses that belong to a website, based on your domain name. It also has a plugin version for *Firefox*. A free account offers 25 searches per month at no cost, ideal for small searches such as those of this course.

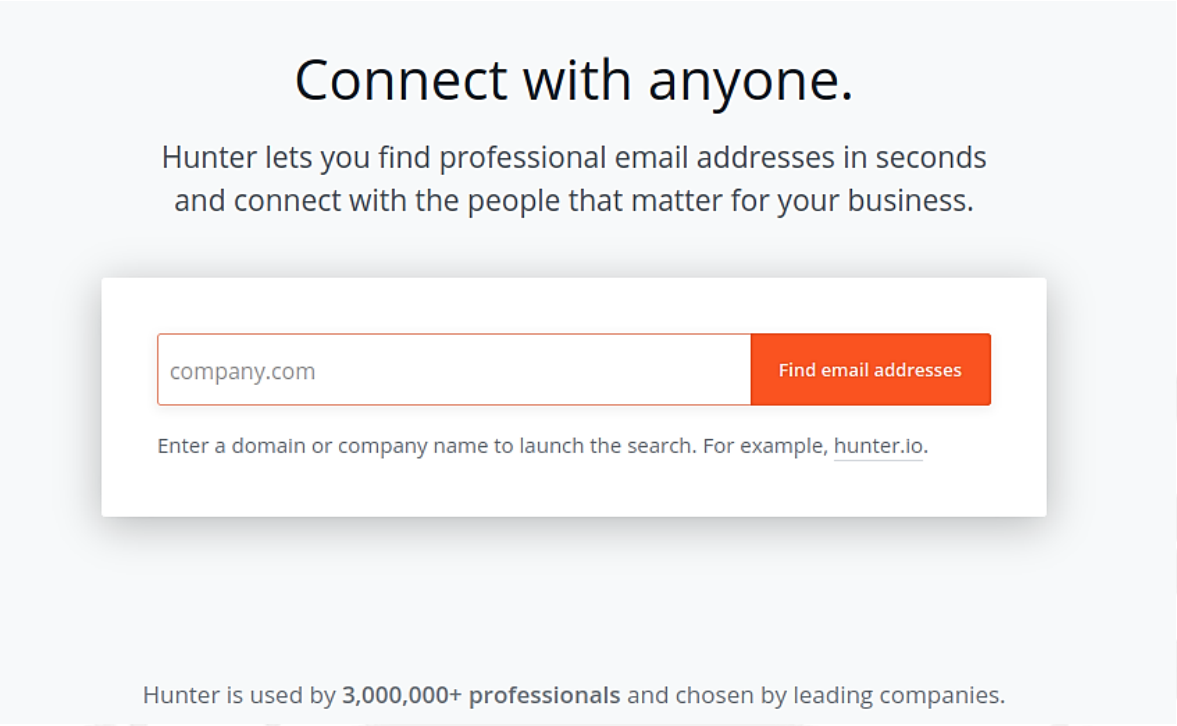


Figure 209. Hunter.io

LeakCheck (https://leakcheck.io/) and Have I been pwnd? (https://haveibeenpwned.com/)

Both are search engines to locate leaked passwords associated with emails in various services. The free use of *LeakCheck* limits the detail of the results and the number of monthly searches, but it is fully operational.

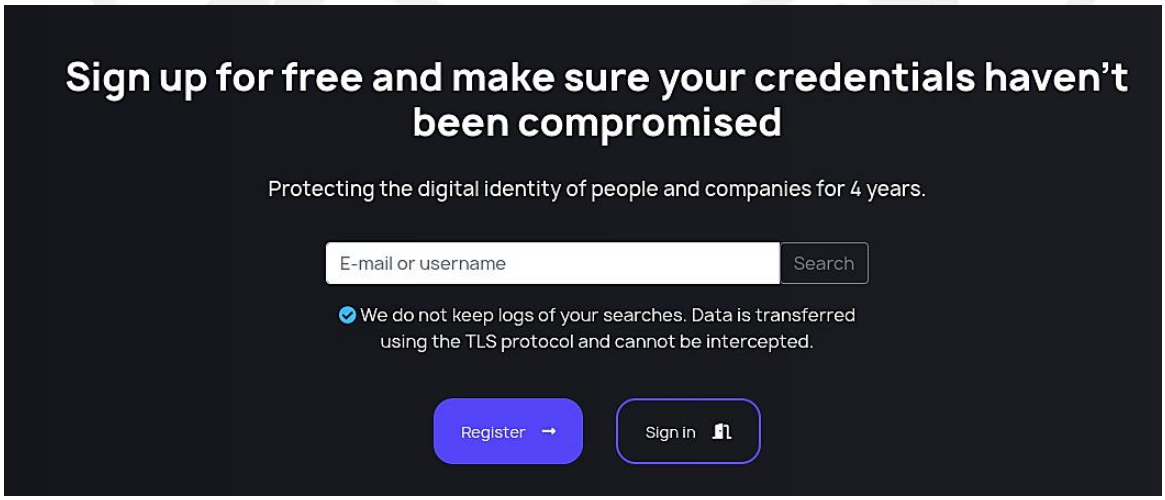


Figure 210. *LeakCheck*

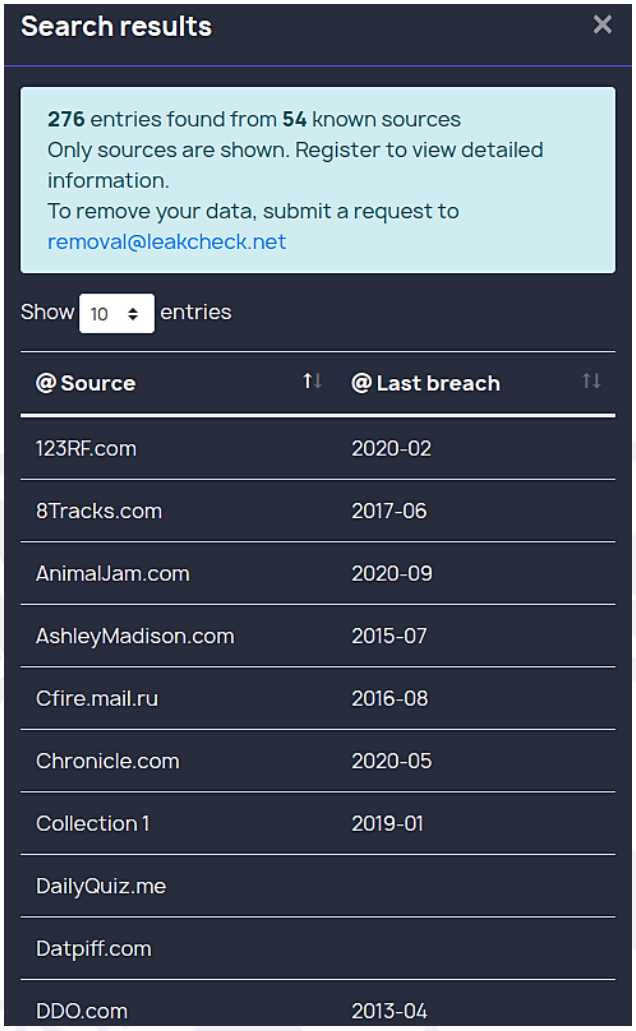


Figure 211. Example of leak sources located by *LeakCheck*

Haveibeenpwnd not only allows that as we saw in **Lab 2**, but also contemplates listing sites that have been compromised, knowing if a specific password has been leaked, if a domain has been compromised (if you verify that it is yours first) and notifying you via email if your address appears in any of the dumps that the web updates frequently.

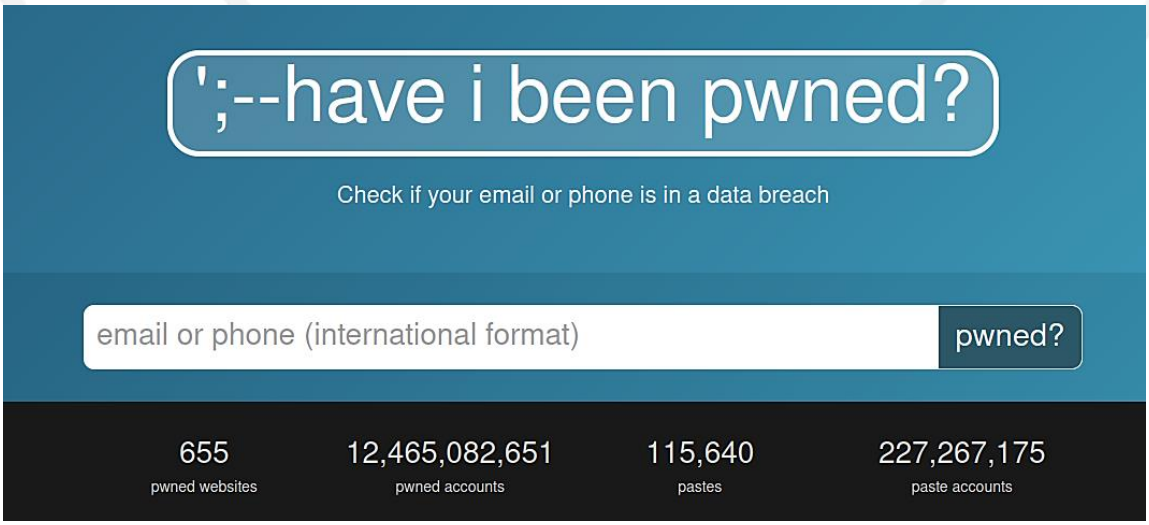


Figure 212. *Have I been pwnd?*

Search engines for source or executable code

Grep.App (https://grep.app/)

It searches half a million Git repositories, supporting regular expressions.

Figure 213. *Grep.app*

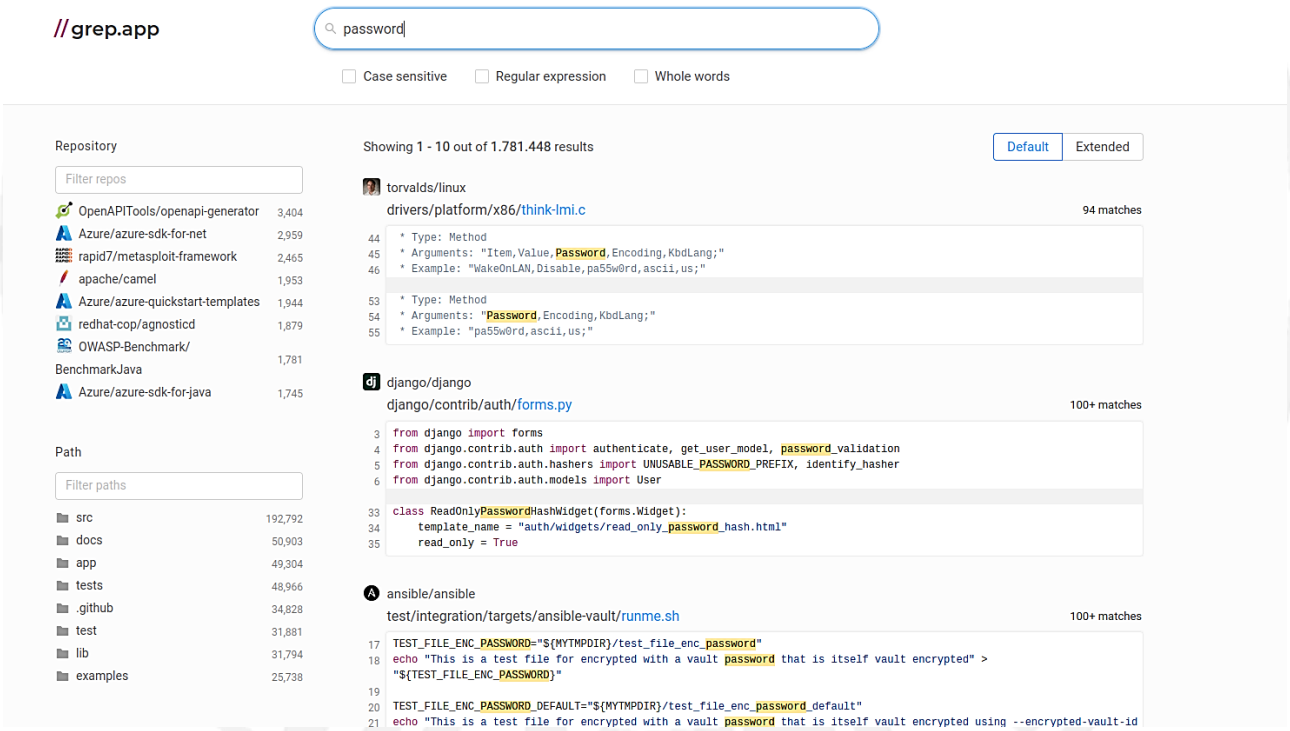


Figure 214. *grep.app* code search system

Joe SandBox (https://www.joesandbox.com/)

Seeing this here may seem quite strange to you, since it does not seem like a tool that can be used to find out anything about a victim, but do not be fooled by the purpose of a ***sandbox*** (run programs and see if they have strange behaviors that may indicate that they are malware); Applied on an executable or document that offers a goal on their websites, for example, part of the work they must do **can reveal interesting data for our investigations**:

- If the file is **signed**, it **can reveal data of its author** to use in an investigation.

- If during the execution **the program contact an external site**, it can be an ideal way to discover domains belonging to the target that may otherwise go unnoticed. Yes, I am talking about detecting "hidden" places from a target that we find because some program that has it connects to them.
- If it turns out that the target is offering malware to its customers, this means two things:
    - Your target **is a criminal**
    - Your target **has been trojanized** and has an infection in progress, which has reached the files it offers to others (*supply-chain attack*). Apart from what this implies at the level of your audit (you will probably have to launch a few alarms ...), the type of infection and the behavior of the executable **can reveal a backdoor in the target**, that is, a site through which the attacker "sneaks" to manipulate the target's systems and... that now you can also use if it serves something in your research. However, this is "engagement" and is not part of the course.

Joe Sandbox is one of these best-known analysis environments and, although it has a rather defensive use, we have just seen how its deep analysis capabilities of the behavior of an executable can be used for something offensive character 😉.

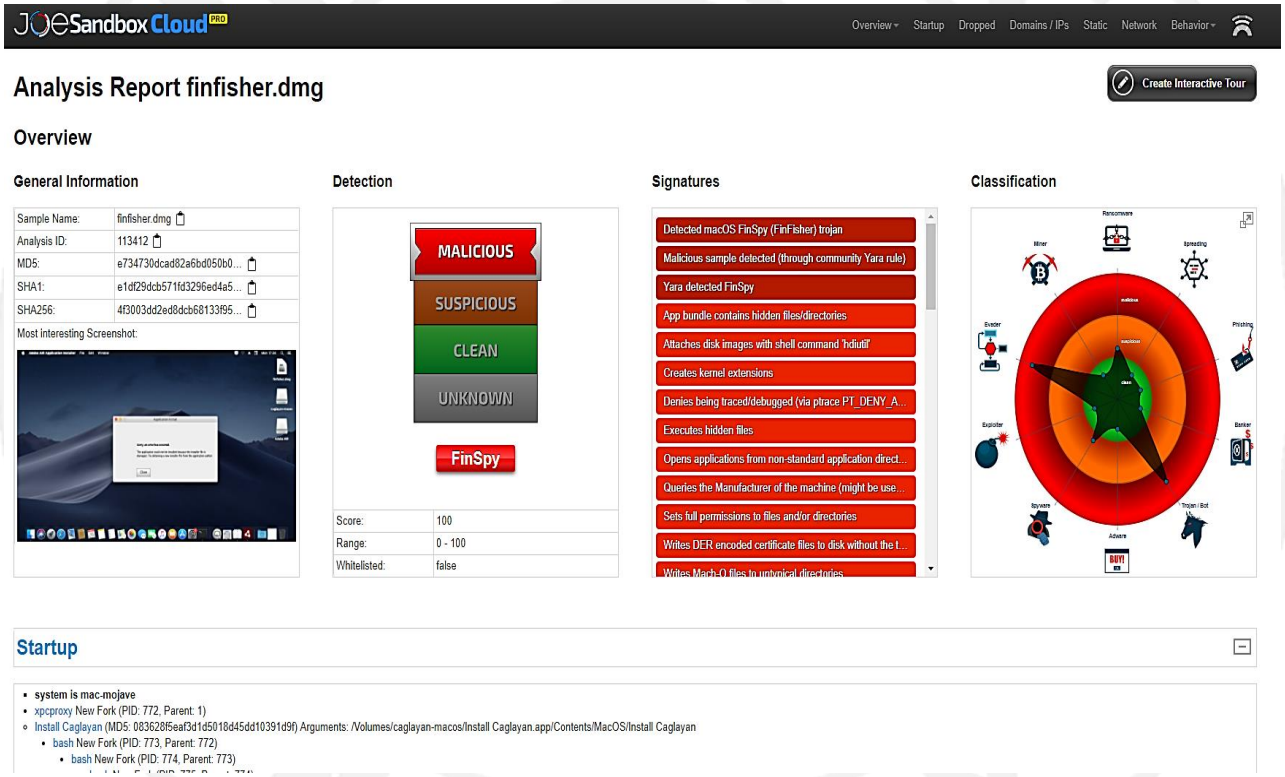


Figure 215. Joe "dissecting" a malware. Source: https://www.joesecurity.org/blog/1896065926254527264

SearchCode (https://searchcode.com/)

Another code finder within repositories, although in this case with a much larger database than other similar solutions on the market.

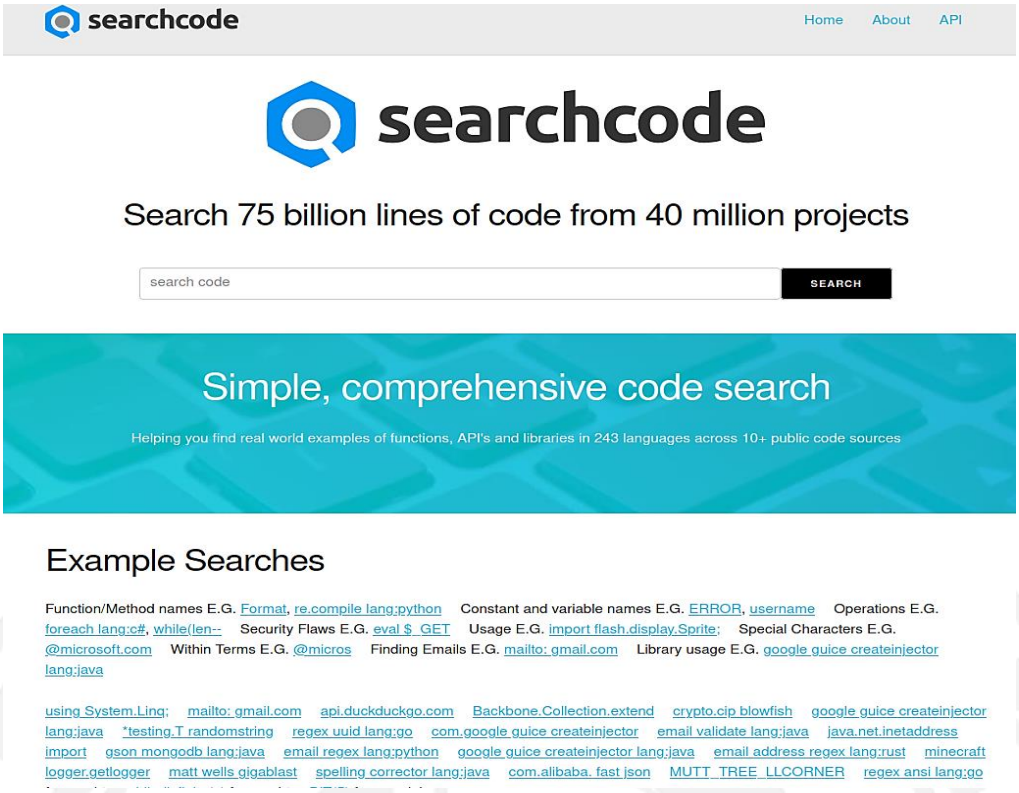


Figure 216. *SearchCode*

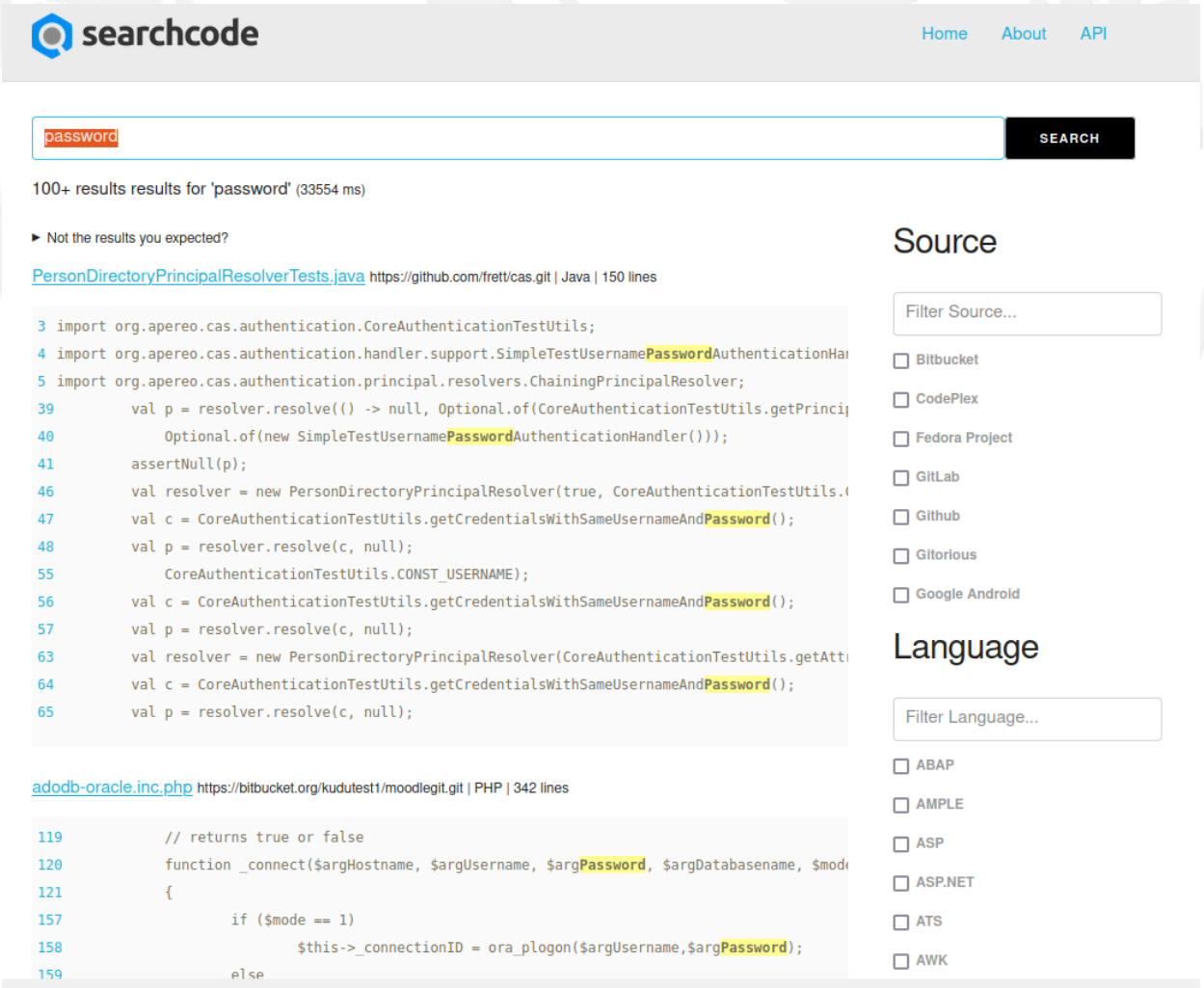
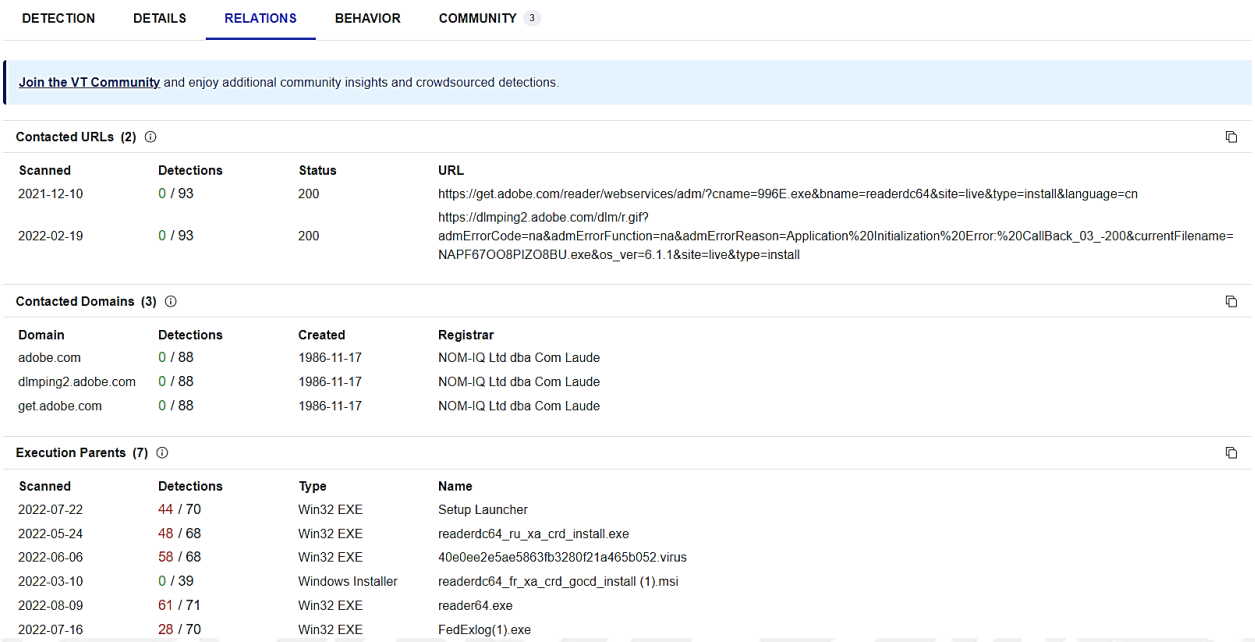


Figure 217. *SearchCode* search example

Virustotal (https://www.virustotal.com/gui/home/upload) and VT Intelligence (https://www.virustotal.com/gui/intelligence-overview)

Virustotal has been recently renewed and **can now be used for the same thing we saw in *Joe Sandbox***: reveal paths and data of an executable without having to run it ourselves: it is the *Virustotal* sandbox that will do it for us, on its own servers. As you can see, its use goes far beyond the basic use we saw in **Laboratory 1**. The following two images show clear examples of this. It is also easier to use than Joe: it does not require any registration! 😊.

The ***VT Intelligence*** platform instead is not free and promises to be something of a worldwide threat intelligence search engine of *Google's* scale, establishing relationships between data as a social network of people. In other words, it is a platform that allows us to analyze global threats and focus into the details only on those that interest us. Unfortunately, being not free means it is outside the objective of this course.



DETECTION | DETAILS | RELATIONS | BEHAVIOR | COMMUNITY 3

Join the VT Community and enjoy additional community insights and crowdsourced detections.

Contacted URLs (2)

Scanned	Detections	Status	URL
2021-12-10	0 / 93	200	https://get.adobe.com/reader/webservices/adm/?cname=996E.exe&bname=readerdc64&site=live&type=install&language=cn
2022-02-19	0 / 93	200	https://dlmping2.adobe.com/dlm/r.gif?admErrorCode=na&admErrorFunction=na&admErrorReason=Application%20Initialization%20Error:%20CallBack_03_-200&currentFilename=NAPF67OO8PIZO8BU.exe&os_ver=6.1.1&site=live&type=install

Contacted Domains (3)

Domain	Detections	Created	Registrar
adobe.com	0 / 88	1986-11-17	NOM-IQ Ltd dba Com Laude
dlmping2.adobe.com	0 / 88	1986-11-17	NOM-IQ Ltd dba Com Laude
get.adobe.com	0 / 88	1986-11-17	NOM-IQ Ltd dba Com Laude

Execution Parents (7)

Scanned	Detections	Type	Name
2022-07-22	44 / 70	Win32 EXE	Setup Launcher
2022-05-24	48 / 68	Win32 EXE	readerdc64_ru_xa_crd_install.exe
2022-06-06	58 / 68	Win32 EXE	40e0ee2e5ae5863fb3280f21a465b052.virus
2022-03-10	0 / 39	Windows Installer	readerdc64_fr_xa_crd_gocd_install (1).msi
2022-08-09	61 / 71	Win32 EXE	reader64.exe
2022-07-16	28 / 70	Win32 EXE	FedExlog(1).exe

Figure 218. *Virustotal* revealing third-party sites by connecting an executable with them

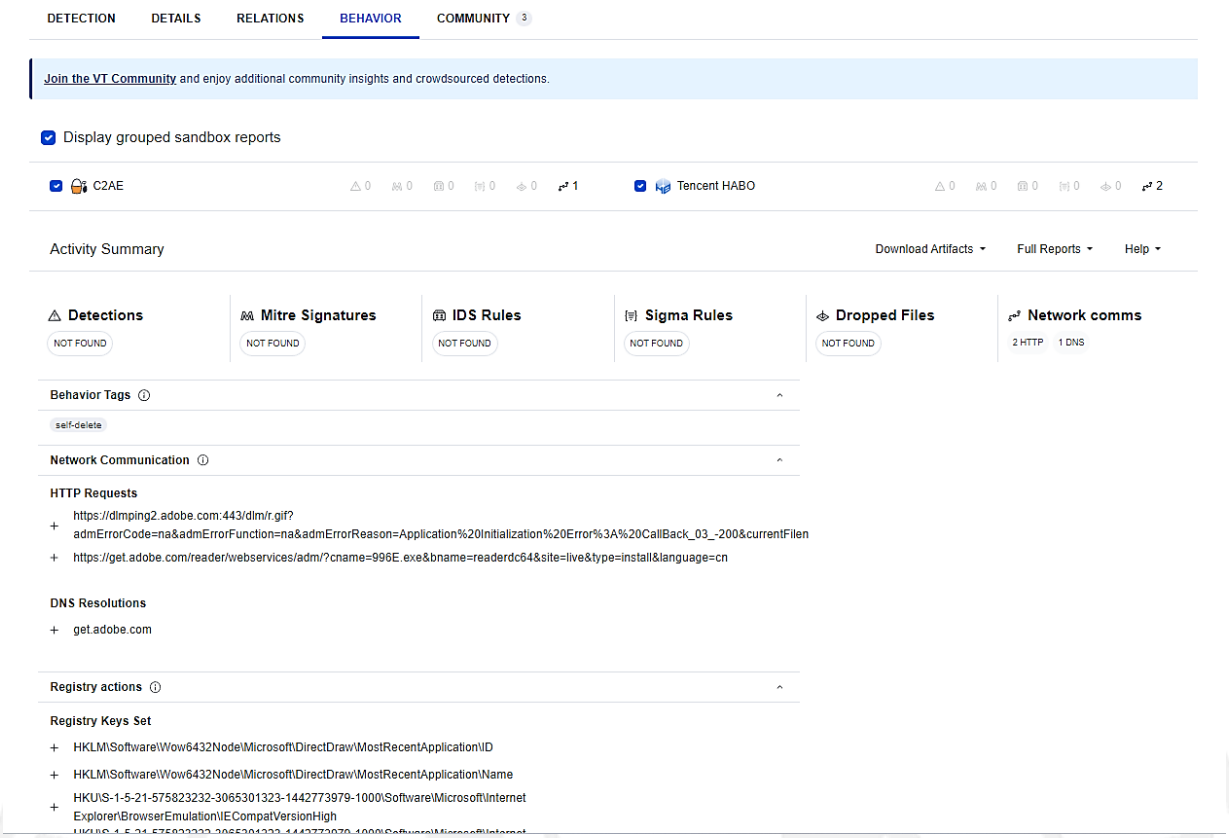


Figure 219. Inspecting the behavior of the *Virustotal* executable can also reveal these external paths, among other things

When the end justifies the means

Phishing and Spear Phishing to obtain information

This chapter may make you think I have gone crazy. Not yet 😊. I just want you to understand that OSINT has an "evil" side, i.e., it can be used to commit crimes. All the information collected through the techniques seen **can be used for different attacks that use social engineering**. And, as you can see in this image, there are quite a few:

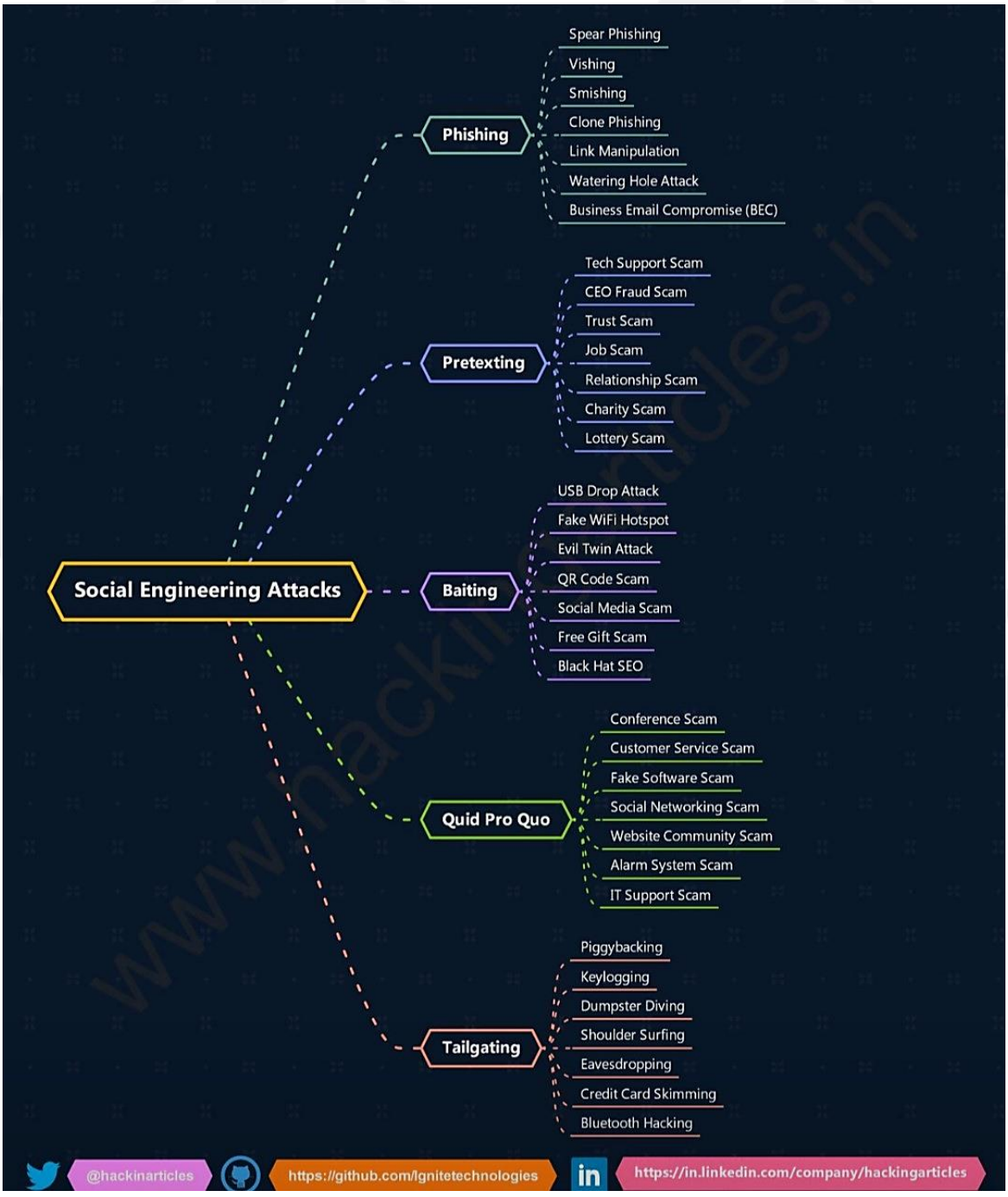


Figure 220. Many people do not understand the severity and importance of social engineering attacks

But in this section we are going to focus on one of the most important, ***phishing* and the tools that facilitate it**, which have a legal use. This usage exists because, although it seems incredible, **many companies make simulations of phishing attacks and spear phishing** as training for their employees. In other words, they use professional phishing tools to launch simulated campaigns, but exactly as criminals would. Then, they extract the information they can from them, and, with that, they make a report of the vulnerability of the company to these attacks[2]. Considering what we said at the beginning of this document about the number of attacks that come from human carelessness, doing this kind of training already makes more sense, *right?* 😊 tag. And, since we make them, we do them well 😉.

In this section we are going to talk about three tools for this purpose, which are ***SET***, ***Gophish*** and ***Evilginx2***. Another similar tool is ***SniperPhish*** (https://github.com/GemGeorge/SniperPhish). On the other hand, although we are going to talk about email communication, doing a scam via SMS is also viable with services like this: https://www.hackers-arise.com/post/social-engineering-attacks-creating-a-fake-sms-message

SET

Introduction

Social-Engineer Toolkit (SET) is an open source pentesting tool designed to make phishing attacks through social engineering. Therefore, the attacks you make can be "fed" by the information you have found out through OSINT about the victim in everything we have described so far.

SET has a set of attack vectors that can be used for the ***TA0001 Initial Access*** and ***TA0002 Execution*** phases of the MITRE ATT&CK, and it is not a tool to be taken as a joke: the attacks it generates **can be perfectly real**. The success or not of them will depend on how good or bad the message you send to the victim is, following the techniques and psychological guidelines we have seen. SET is installed by default on *Kali Linux*, so we simply have to open it.

[2] Maybe you think: *What happens if I take the bait? Do they fire me?* Well, everything is possible, but for now my experience is that you will be forced to take an education course about phishing attacks. Not very punishing 😊

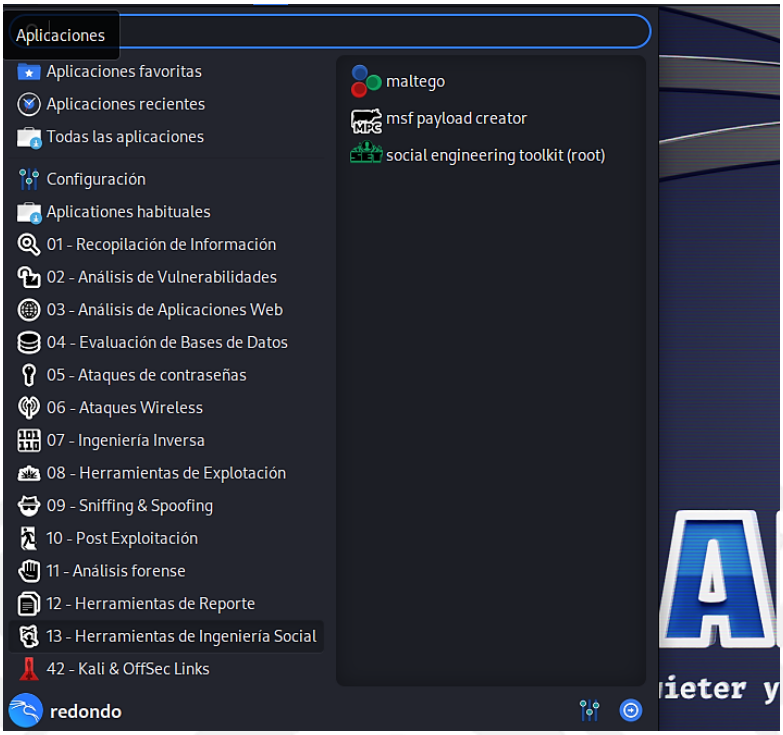


Figure 221. Opening SET on *Kali Linux*

As I said before, the possibility of doing real damage with this tool exists and is not negligible, so using it you admit knowing the risks. The tool was created to check a victim's vulnerability to phishing and social engineering attacks, but as you can imagine, doing this can easily become a crime if we do not know when to stop (which the contract we have signed will tell us).

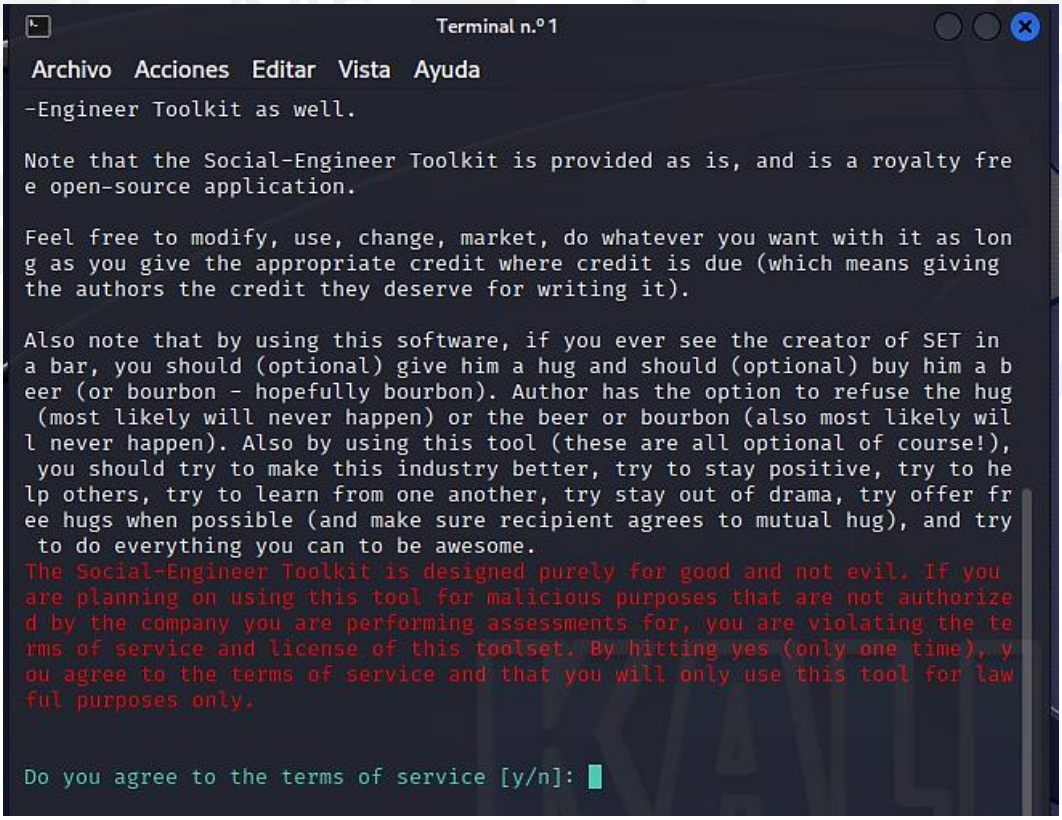


Figure 222. SET Liability Clause

SET is a command-line tool **based on menus** that are selected by number, in the style of the programs one does in the first year of the degree. Each option opens the options to make a different type of attack, but in this section we are interested in first doing one about **social engineering**.

Figure 223. SET Main Menu

Phishing emails with SET

Within the first option of the menu, we have the possibility to **create spear phishing messages**. This allows us to generate messages of this style and send them to a victim with payloads attached. For example, you can send a malicious PDF so that when the victim opens it, it compromises the system. It also allows you to easily spoof the sending address if you also have the `sendmail` package installed (`apt-get install sendmail`) and change in `config/set_config` the entry `SENDMAIL=OFF` to `SENDMAIL=ON`. The following image shows the submenu that appears when we choose the option of social engineering attacks.

Figure 224. Social engineering attacks available in SET

Although so far the thing seems very easy, you have to be realistic: If you send the mail yourself to the victims, and they have the typical mail provider of large company in the cloud (*Outlook*, *Gmail* ...) the chances that the mail do arrive or arrives, but it is not classified as spam **are very low**. Nowadays mail services are an oligopoly and large companies have in place protections in the mail (SPF, DKIM, DMARC) that **will detect that the sender is not reliable** or that it directly comes with malware. The same happens if we copy and paste the mail and try to send it through an account of one of these providers, ours or "acquired" (***TA0042 Resource Development***). If you have a malicious payload, the most normal thing is that the provider detects it.

However, this does not mean that it will always fail. There are still victims who have their own mail system (usually very poorly protected), and the payload may not be detected as malware (especially if it is generated with `msfvenom` **and encoded** (**Topic 7**). There is also the option to send a link in the mail, pointing to a copy of a real website controlled by us.

There are two ways to do a spear phishing attack:

1. Make a massive attack on emails.
2. Create a payload and a template for social engineering.

For the first **SET does all the work for us** (option 1). In the case of the second, we must create a payload and use it in an attack that we configure. They correspond to options 1 and 2 of this menu. In the following sequence of images, you can see the successive menu options to choose to make an attack of this class. We see that the sequence of steps to follow is aligned with what was explained in **laboratories 11** and **12** of the course.

```
The Spearphishing module allows you to specially craft email messages and send
them to a large (or small) number of people with attached fileformat malicious
payloads. If you want to spoof your email address, be sure "Sendmail" is in-
stalled (apt-get install sendmail) and change the config/set_config SENDMAIL=OFF
flag to SENDMAIL=ON.

There are two options, one is getting your feet wet and letting SET do
everything for you (option 1), the second is to create your own FileFormat
payload and use it in your own attack. Either way, good luck and enjoy!

  1) Perform a Mass Email Attack
  2) Create a FileFormat Payload
  3) Create a Social-Engineering Template

 99) Return to Main Menu

set:phishing>
```

Figure 225. Types of spear phishing attack

```
set:phishing>1
/usr/share/metasploit-framework/

 Select the file format exploit you want.
 The default is the PDF embedded EXE.

           ********** PAYLOADS **********

   1) SET Custom Written DLL Hijacking Attack Vector (RAR, ZIP)
   2) SET Custom Written Document UNC LM SMB Capture Attack
   3) MS15-100 Microsoft Windows Media Center MCL Vulnerability
   4) MS14-017 Microsoft Word RTF Object Confusion (2014-04-01)
   5) Microsoft Windows CreateSizedDIBSECTION Stack Buffer Overflow
   6) Microsoft Word RTF pFragments Stack Buffer Overflow (MS10-087)
   7) Adobe Flash Player "Button" Remote Code Execution
   8) Adobe CoolType SING Table "uniqueName" Overflow
   9) Adobe Flash Player "newfunction" Invalid Pointer Use
  10) Adobe Collab.collectEmailInfo Buffer Overflow
  11) Adobe Collab.getIcon Buffer Overflow
  12) Adobe JBIG2Decode Memory Corruption Exploit
  13) Adobe PDF Embedded EXE Social Engineering
  14) Adobe util.printf() Buffer Overflow
  15) Custom EXE to VBA (sent via RAR) (RAR required)
  16) Adobe U3D CLODProgressiveMeshDeclaration Array Overrun
  17) Adobe PDF Embedded EXE Social Engineering (NOJS)
  18) Foxit PDF Reader v4.1.1 Title Stack Buffer Overflow
  19) Apple QuickTime PICT PnSize Buffer Overflow
  20) Nuance PDF Reader v6.0 Launch Stack Buffer Overflow
  21) Adobe Reader u3D Memory Corruption Vulnerability
  22) MSCOMCTL ActiveX Buffer Overflow (ms12-027)

set:payloads>
```

Figure 226. Possible payloads of phishing mail to create

```
[-] Default payload creation selected. SET will generate a normal PDF with embedded
 EXE.

    1. Use your own PDF for attack
    2. Use built-in BLANK PDF for attack

set:payloads>
```

Figure 227. Creating a PDF payload with an executable inside, either from an existing PDF or a blank one

```
   1) Windows Reverse TCP Shell               Spawn a command shell on victim and se
nd back to attacker
   2) Windows Meterpreter Reverse_TCP         Spawn a meterpreter shell on victim an
d send back to attacker
   3) Windows Reverse VNC DLL                 Spawn a VNC server on victim and send
back to attacker
   4) Windows Reverse TCP Shell (x64)         Windows X64 Command Shell, Reverse TCP
 Inline
   5) Windows Meterpreter Reverse_TCP (X64)   Connect back to the attacker (Windows
x64), Meterpreter
   6) Windows Shell Bind_TCP (X64)            Execute payload and create an acceptin
g port on remote system
   7) Windows Meterpreter Reverse HTTPS       Tunnel communication over HTTP using S
SL and use Meterpreter

set:payloads>
```

Figure 228. Effect of the payload to be generated. Note how they are all reverse shells (Topic 7)

```
set:payloads>2
set> IP address or URL (www.ex.com) for the payload listener (LHOST) [10.0.2.15]:
set:payloads> Port to connect back on [443]:
[-] Defaulting to port 443...
[!] Metasploit has not been previously run on the system. This means that the msf d
irectories haven't been created yet. Running Metasploit for you.
[*] Waiting 10 seconds for the directories to be created...
```

Figure 229. Configuring the IP of the server that will listen for connection attempts by victims who have opened the PDF

```
[-] As an added bonus, use the file-format creator in SET to create your attachment
.

   Right now the attachment will be imported with filename of 'template.whatever'

   Do you want to rename the file?

   example Enter the new filename: moo.pdf

    1. Keep the filename, I don't care.
    2. Rename the file, I want to be cool.

set:phishing>
```

Figure 230. End the PDF with "bomb" created

At this point SET allows us (as with the menu option called "Mass Mailer Attack") **to send emails to victims** automatically with a personalized message that we can elaborate from the OSINT information we have collected. You have two options:

- *E-Mail Attack Single Email Address*
- *E-Mail Attack Mass Mailer*

The first allows you to send an email to a single email address (typical **spear phishing**), while the second would do it to a list of recipients that we previously import, as wide as we want. With this option you can make a **targeted attack against a group of people**, for example, since the message will be sent to all members of this list.

```
   Social Engineer Toolkit Mass E-Mailer

   There are two options on the mass e-mailer, the first would
   be to send an email to one individual person. The second option
   will allow you to import a list and send it to as many people as
   you want within that list.

   What do you want to do:

   1.  E-Mail Attack Single Email Address
   2.  E-Mail Attack Mass Mailer

   99. Return to main menu.

set:phishing>
```

Figure 231. Configuration of the type of sending emails to be made

```
   Do you want to use a predefined template or craft
   a one time email template.

   1. Pre-Defined Template
   2. One-Time Use Email Template

set:phishing>
```

Figure 232. Composition of phishing email

In this precise moment **you can compose your own email template** or use a predefined one. The second option is just for quick tests, as these default templates are known and the chances to be identified by defensive software are very high. This is the moment in which **you should put into play all the information you have collected from the target** to compose your own custom email, increasing the **credibility** (and hence, the **effectiveness**) of the attack.

```
set:phishing>1
[-] Available templates:
1: New Update
2: How long has it been?
3: Have you seen this?
4: Status Report
5: Dan Brown's Angels & Demons
6: Strange internet usage from your computer
7: Computer Issue
8: Baby Pics
9: WOAAAA!!!!!!!!!! This is crazy ...
10: Order Confirmation
set:phishing>10
set:phishing> Send email to:
```

Figure 233. Predefined phishing emails

```
  1. Use a gmail Account for your email attack.
  2. Use your own server or open relay

set:phishing>
```

Figure 234. Choosing the type of server through which the emails will be sent

```
set:phishing>2
set:phishing> From address (ex: moo@example.com):
set:phishing> The FROM NAME user will see:
set:phishing> Username for open-relay [blank]:
Password for open-relay [blank]:
set:phishing> SMTP email server address (ex. smtp.youremailserveryouown.com):
set:phishing> Port number for the SMTP server [25]:
set:phishing> Flag this message/s as high priority? [yes|no]:
[!] valid responses are 'n|y|N|Y|no|yes|No|Yes|NO|YES'
set:phishing> Flag this message/s as high priority? [yes|no]:no
[*] Unable to connect to mail server. Try again (Internet issues?)
[*] SET has finished delivering the emails
set:phishing> Setup a listener [yes|no]:
```

Figure 235. Email sending process (not working, because we have not set up the server on purpose)

Web phishing attacks with SET

If we go to the previous menu and select "***The web attack vectors***", we will access a simple way to use web-based attacks that can compromise the victim. Basically, they are used to create phishing attacks that compromise anyone who clicks on a link that generates one of these options, such as those shown in this image. Realize that things like "***Credential harverster***" are more suitable for what we are describing in this document.

```
The Web Attack module is a unique way of utilizing multiple web-based attacks in or
der to compromise the intended victim.

The Java Applet Attack method will spoof a Java Certificate and deliver a metasploi
t based payload. Uses a customized java applet created by Thomas Werth to deliver t
he payload.

The Metasploit Browser Exploit method will utilize select Metasploit browser exploi
ts through an iframe and deliver a Metasploit payload.

The Credential Harvester method will utilize web cloning of a web- site that has a
username and password field and harvest all the information posted to the website.

The TabNabbing method will wait for a user to move to a different tab, then refresh
 the page to something different.

The Web-Jacking Attack method was introduced by white_sheep, emgent. This method ut
ilizes iframe replacements to make the highlighted URL link to appear legitimate ho
wever when clicked a window pops up then is replaced with the malicious link. You c
an edit the link replacement settings in the set_config if its too slow/fast.

The Multi-Attack method will add a combination of attacks through the web attack me
nu. For example you can utilize the Java Applet, Metasploit Browser, Credential Har
vester/Tabnabbing all at once to see which is successful.

The HTA Attack method will allow you to clone a site and perform powershell injecti
on through HTA files which can be used for Windows-based powershell exploitation th
rough the browser.

   1) Java Applet Attack Method
   2) Metasploit Browser Exploit Method
   3) Credential Harvester Attack Method
   4) Tabnabbing Attack Method
   5) Web Jacking Attack Method
   6) Multi-Attack Web Method
   7) HTA Attack Method

  99) Return to Main Menu

set:webattack>
```

Figure 236. Types of web attacks (and their description) available in SET

On the next menu we clearly see that we have the option of using a predefined web template, our own one or just **clone a real website**, which can be very effective with little effort.

```
set:webattack>2

 The first method will allow SET to import a list of pre-defined web
 applications that it can utilize within the attack.

 The second method will completely clone a website of your choosing
 and allow you to utilize the attack vectors within the completely
 same web application you were attempting to clone.

 The third method allows you to import your own website, note that you
 should only have an index.html when using the import website
 functionality.

   1) Web Templates
   2) Site Cloner
   3) Custom Import

  99) Return to Webattack Menu

set:webattack>
```

Figure 237. Obtaining the fake website to trigger the attack

Media infected with SET

In the main menu we also have the module "***Infectious USB/CD/DVD***" that will create an `autorun.inf` file and a `metasploit` payload (**Topic 7**, **Laboratories 11 and 12**). If we copy both things to a DVD / USB and the victim inserts that media into his machine, he will activate the auto-execution option and this payload will be executed, compromising the system. To generate the payload, you can use these two options.

- *File-Format Exploits*
- *Standard Metasploit Executable*

```
 The Infectious USB/CD/DVD module will create an autorun.inf file and a
 Metasploit payload. When the DVD/USB/CD is inserted, it will automatically
 run if autorun is enabled.

 Pick the attack vector you wish to use: fileformat bugs or a straight executable.

   1) File-Format Exploits
   2) Standard Metasploit Executable

  99) Return to Main Menu

set:infectious>
```

Figure 238. Payload types for infected media on SET

Payloads and listeners

The "***Create payload and listener***" option in the main menu is a simple way to create a *Metasploit* payload, much more than doing it by hand (as explained in **Topic 7**). This option will export the executable that we have to send to the victim (and convince him to execute it) and a listener for him. Any victim who executes it will send a request to our listener, and with it we can make a ***reverse shell***. In other words, the same sequence of events that we have seen in the execution part of an attack chain, but with this option we will have the payload created **to use where we think it is appropriate**.

```
   1) Windows Shell Reverse_TCP               Spawn a command shell on victim and s
end back to attacker
   2) Windows Reverse_TCP Meterpreter         Spawn a meterpreter shell on victim a
nd send back to attacker
   3) Windows Reverse_TCP VNC DLL             Spawn a VNC server on victim and send
 back to attacker
   4) Windows Shell Reverse_TCP X64           Windows X64 Command Shell, Reverse TC
P Inline
   5) Windows Meterpreter Reverse_TCP X64     Connect back to the attacker (Windows
 x64), Meterpreter
   6) Windows Meterpreter Egress Buster       Spawn a meterpreter shell and find a
port home via multiple ports
   7) Windows Meterpreter Reverse HTTPS       Tunnel communication over HTTP using
SSL and use Meterpreter
   8) Windows Meterpreter Reverse DNS         Use a hostname instead of an IP addre
ss and use Reverse Meterpreter
   9) Download/Run your Own Executable        Downloads an executable and runs it

set:payloads>
```

Figure 239. Payloads (and their listeners) created with SET

Powershell

Finally, if we return to the previous menu and press "9" we go to "***Powershell Attack Vector***" where we will have these options.

```
set> 9

The Powershell Attack Vector module allows you to create PowerShell specific attack
s. These attacks will allow you to use PowerShell which is available by default in
all operating systems Windows Vista and above. PowerShell provides a fruitful  land
scape for deploying payloads and performing functions that  do not get triggered by
 preventative technologies.

   1) Powershell Alphanumeric Shellcode Injector
   2) Powershell Reverse Shell
   3) Powershell Bind Shell
   4) Powershell Dump SAM Database

  99) Return to Main Menu

set:powershell>
```

Figure 240. PowerShell attacks available with SET

This module allows you to create specific attacks with *PowerShell*, which is installed by default on Windows operating systems and some Linux. This attack vector has a large number of payloads with functions that can easily not be detected by modern antivirus technologies, although it is beyond the scope of this course.

- *PowerShell Alphanumeric Shellcode Injector*
- *PowerShell Reverse Shell*
- *PowerShell Bind Shell*
- *PowerShell Dump SAM Database*

Gophish

Gophish (https://getgophish.com/) is a **phishing framework** that **makes simulating a real phishing attack a breeze**. The idea is simply that anyone can launch one of these attacks in order to train their employees quickly and without problem. It is free and widely used in companies that are dedicated to security. The reason is simple: it allows you to launch separate campaigns, total configuration of the emails to be sent, choose the type of payload, manage in real time all the "victims" of the phishing emails launched, and other management tools that make it very advanced, all via the web.

However, to start it up **we would need an IP contactable by the machines of the possible phishing victims** (usually public) and a mail server or account through which to send phishing emails in an automated way (the tool would do it through which we configure it). Due to the extra difficulty that this entails, and that obtaining an adequate testing infrastructure can be an expense for you (cloud servers), we are not going to include it in the course. However, mentioning it so that you know its existence is mandatory due to the use it has in private companies, and therefore you could have the option of orchestrating these campaigns in your work. If you think you have the necessary testing infrastructure, the installation and commissioning is done very easily by following the steps in this link: https://docs.getgophish.com/user-guide/

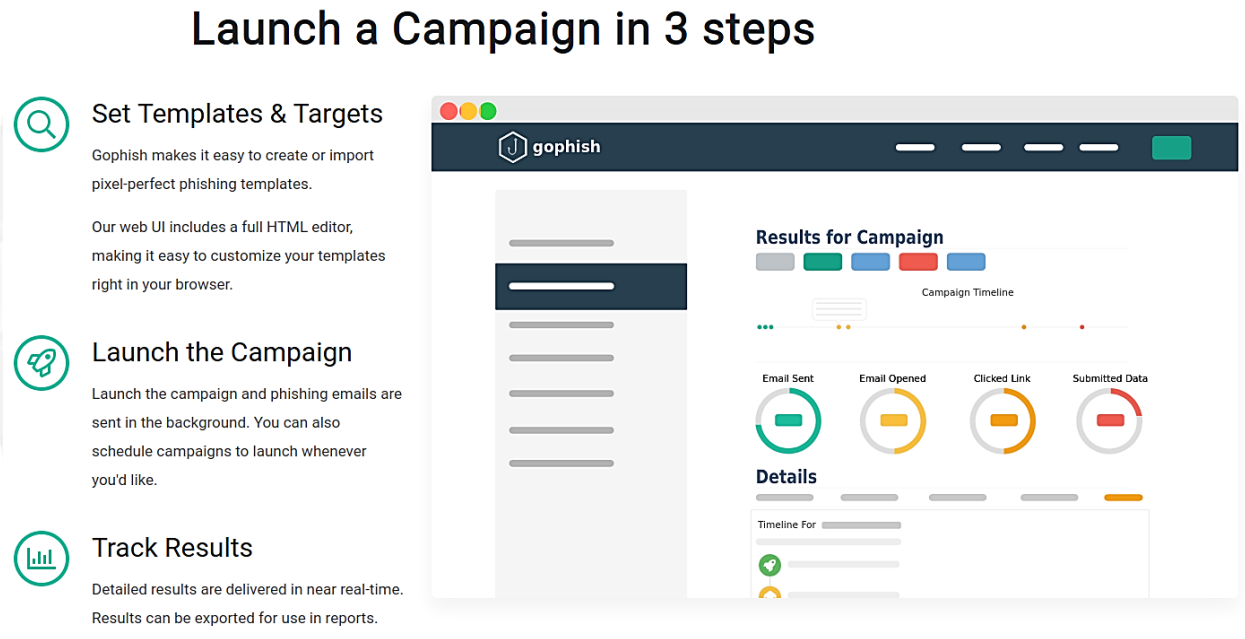


Figure 241. ***GoPhish***

Evilgnx2

Evilgnx2 (https://github.com/kgretzky/evilginx2) is a framework **for creating *Attacker-in-the-Middle* (AitM) attacks** with the aim of obtaining **user credentials and their session cookies** through phishing attacks. Obtaining cookies also has the consequence of being able to **bypass advanced security measures, such as 2FA**. The use of this tool requires the same thing as *Gophish*: an IP contactable by victims, usually on the Internet. That is why we will not include it in this course either, although if you have the possibility of having that infrastructure, you can follow this tutorial to start it and test it.

- https://thehackerway.com/2019/11/06/phishing-2-0-con-evilginx2-parte-1/
- https://thehackerway.com/2020/01/02/phishing-2-0-con-evilginx2-parte-2/

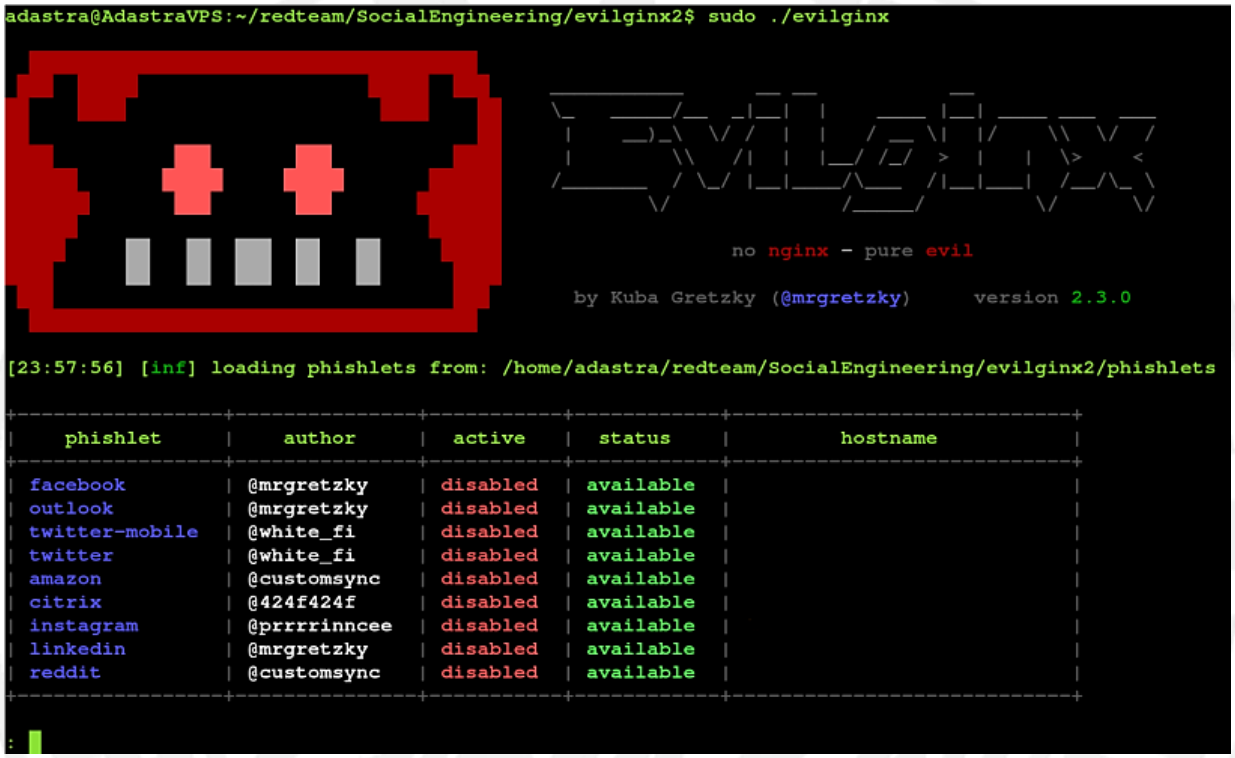


Figure 242. *EvilGinx2*

Dark Web Investigations

I am aware that this section of the documentation is controversial because you may think (and it really is for the most part) that on the *dark web* there are only **crimes and information to commit them**. This information **is not given for you to do as part of the work**, but to make you realize that it is really important for companies and organizations to monitor the *dark web* for several reasons:

- **Cyberattack prevention**: The *dark web* is a place where cybercriminals can share and sell stolen or leaked information, such as compromised passwords, credentials, intellectual property, and other sensitive data. If a company or organization monitors the *dark web,* **it can detect if its information has been compromised** and take action to respond to the problem.

  In an OSINT investigation, finding compromised private information can give "the keys to the kingdom" to a criminal. For all these reasons, keeping an eye on the *dark web* (or having a contracted service that does it for you if you are a company), is a necessity, because if instead of doing it

yourself a criminal finds your information, they can enter your infrastructure simply with data that they have found "out there".

- **Rapid incident response**: Without well-done monitoring of the *dark web*, a business or organization **can go months without knowing that a data breach has occurred**. By the time they want to realize these cases, it will be too late.
- **Protection against financial loss**: The financial consequences of a data breach can cost an organization or business a great deal of money, which can be lessened if they know exactly what data has been compromised and, based on them, which affected systems need to be fixed.
- **Legal compliance**: Depending on the nature of the data that has been breached, organizations may be subject to legal action from regulatory bodies, **including fines if negligence is proven** (e.g., the AEPD). Detecting the leak early and remedying the problem as soon as possible can save a lot of money on fines.
- **Reputational damage**: The loss of trust caused by an attack may lead to the loss of customers and also create difficulties in winning new ones. But if the extent of the leak is known, an **appropriate communication policy** can be developed to try to minimize these impacts.

Monitoring the web requires experience and the right tools. For something basic, you can use a connection from a virtual machine with a ***ToR Browser*** in strict mode (**Lab 7**) to avoid as many problems as possible when browsing. You would then only need the `.onion` links from *dark web sites* where the leaked data is dumped, something that is not very difficult to find .

These links may help you. However, I insist that even if this is not *engagement*, **it has a non-negligible risk and is only given for your information**, not being necessary to do it as part of the job:

- https://www.fortinet.com/resources/cyberglossary/dark-web-monitoring
- https://webz.io/blog/dark-web-monitoring/dark-web-monitoring-your-essential-guide/
- https://blog.elhacker.net/2023/04/investigaciones-osint-en-la-dark-web.html
- https://osintteam.blog/dark-web-how-to-get-started-658409184c06?gi=b37e78471105

HACK FIRST • HACK HARD
大和
Cobra Kai
• NO MERCY •

Addendum: A friend wants to know more about these topics (okay, the friend is me): Sites where you can learn more about OSINT

As we said in the introduction, this document is an introduction to this security discipline and although we have tried to cover a lot of common use cases, there is much more to explore. That is why there are plans **to develop this documentation into a full course in the future**. If you want to know more about OSINT while that is happening, I can recommend the following links.

Guides & Articles

- **Introductory article *"A Guide To Open-Source Intelligence (OSINT)":*** https://itsec.group/blog-post-osint-guide-part-1.html
- ***"A Beginners Guide to OSINT"* guide:** https://www.csnp.org/post/a-beginners-guide-to-osint
- ***Telegram OSINT: The Ultimate Guide to Unveiling Insights Beyond What Meets the Eye***: https://espysys.com/blog/telegram-osint-the-ultimate-guide-to-unveiling-insights-beyond-what-meets-the-eye/

Documentation repositories, guides, and tools

- **"Awesome Social Engineering" repository**, with a large number of materials related to how to obtain information of different kinds: https://github.com/giuliacassara/awesome-social-engineering
- **"Awesome OSINT" repository**, with a large number of materials related to how to obtain information of different kinds**:** https://github.com/jivoi/awesome-osint
- **"OSINT Anonymous" repository**, specialized in tools to create false identities complementary to those seen in the document**:** https://github.com/CScorza/OSINTAnonymous
- **"Image Research OSINT" repository**, specialized in tools to carry out various types of research on images: https://github.com/cqcore/Image-Research-OSINT
- **Repository of the course "*Python for OSINT in 21 days*"**, with the materials of this introductory course to OSINT: https://github.com/cipher387/python-for-OSINT-21-days
- **Blog "*OH SHINT!*",** with a huge amount of information on how to get public data on a lot of different topics: https://ohshint.gitbook.io/oh-shint-its-a-blog/
- **"OSINT" repository**, with a large number of materials related to how to obtain information of different kinds: https://github.com/sinwindie/OSINT
- ***OSINT4ALL***: A massive guide to tools adapted to any need we may have in an OSINT investigation: https://start.me/p/L1rEYQ/osint4all?locale=es