OSINT: How to find information on anyone

Your data is more exposed than you think

Petro Cherkasets Follow
May 7 · 20 min read

Open Source Intelligence (OSINT) — is information gathering from publicly available sources and its analysis to produce an actionable intelligence. The scope of OSINT is not limited to cybersecurity only but corporate, business and military intelligence or other fields where information matters.

Whether you are a recruiter, marketing manager, cybersecurity engineer or just a curious person reading the article, you will find something useful for yourself. Maybe you want to know what data of yours is out there for others to find or just want to see if the person or the organization that contacted you online is legit. In this article, I will explain how to discover a person's **digital footprint**, perform **digital investigations**, and gather information for **competitive intelligence** or **penetration testing**.

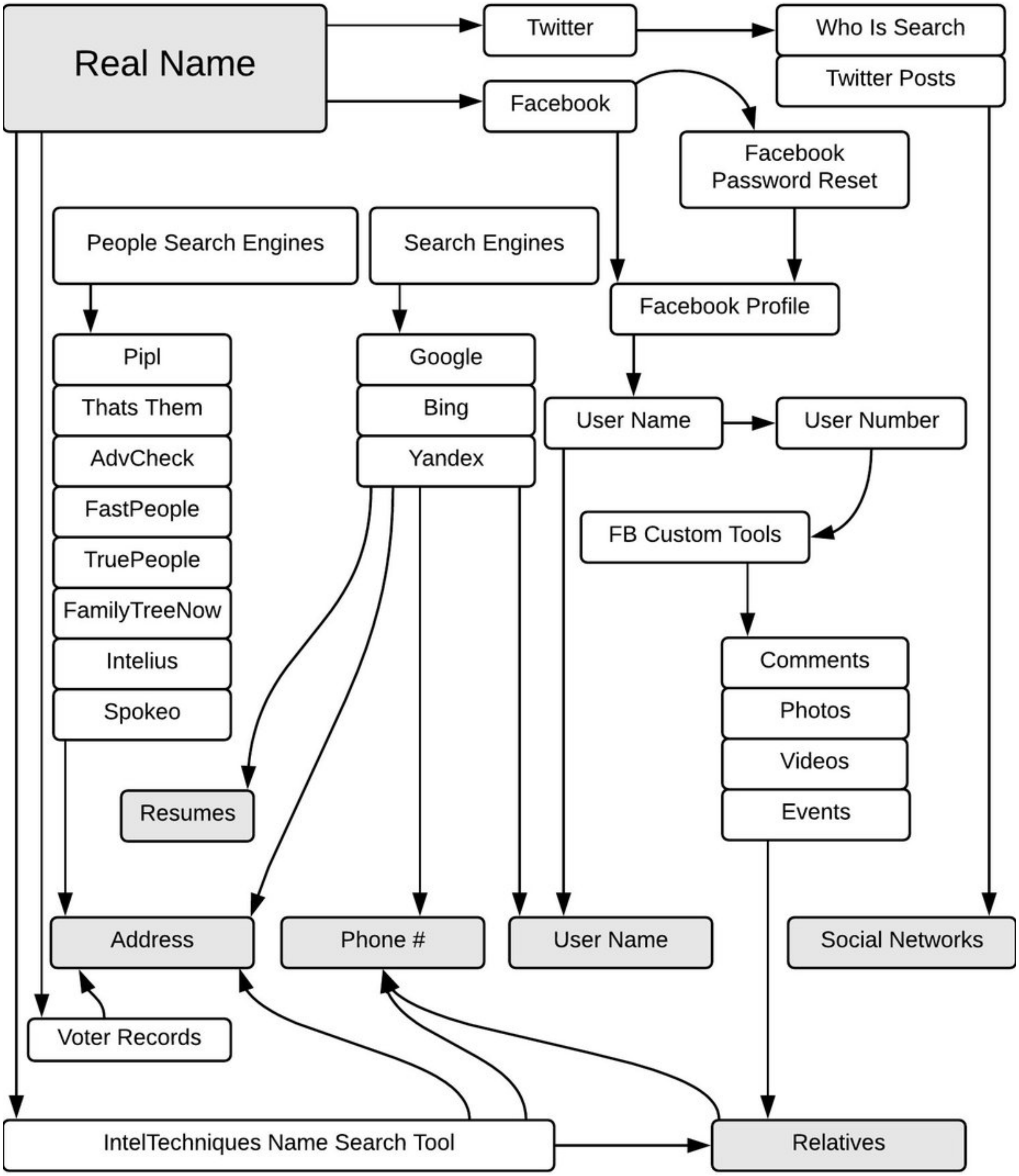
Many OSINT tools are available nowadays so I'm not going to cover them all, only the most popular ones and those useful in the described use cases. In this guide, I show a general approach and different tools and methods that you can use depending on the requirements and the initial data you have.

OSINT steps

1. Start with what you know (email, username, etc.)
2. Define requirements (what you want to get)
3. Gather the data
4. Analyze collected data

5. Pivot as-needed using new gathered data
6. Validate assumptions
7. Generate report

Real name



IntelTechniques.com OSINT Workflow Chart: Real Name

Governmental resources

There are dozens of websites where you can find information about people or organizations and depending on the country, information openness can be different. I'm not going to write about it in details as the governmental resources I would provide might not be relevant to you, as a resident of a different country. Just remember that such resources exist and Google them in need, as they are not that hard to find, especially using the advanced search queries I describe below.

Google Dorks

In 2002, Johnny Long began collecting Google search queries that uncovered vulnerable systems or sensitive information disclosures. He labeled them Google Dorks. Since the article is about legally obtained information I'm not going to show how to get an unauthorized access, however, you can explore Google Hacking Database with thousands of different queries. The queries below can return information that is difficult to locate through a simple search.

- *"john doe" site:instagram.com* — quotation marks force Google Search to do absolutely exact match while the search is performed on Instagram.
- *"john doe" -"site:instagram.com/johndoe" site:instagram.com* — hide postings from the target's own account, but show posted comments on the Instagram posts of others.
- *"john" "doe" -site:instagram.com* — show results that exactly match the given name and surname but in different combinations.

Also, exclude Instagram from results.

- *"CV" OR "Curriculum Vitae" filetype:PDF "john" "doe"* — search for the target's resumes that contain "CV" or "Curriculum Vitae" in the name and have a PDF extension.

Wrap single words in quotes if you are 100% sure about spelling as by default Google will try to shape your keyword to what the masses want. By the way, what's interesting about Instagram is with the right Google Dork you can see comments and likes of private accounts.

Perform a search using advanced search queries on Bing, Yandex, and DuckDuckGo as other search engines might give you results that Google couldn't.

People search

There are websites that specialize in people search which can be done providing a real name, username, email or phone number.

- https://www.spokeo.com
- https://thatsthem.com
- https://www.beenverified.com
- https://www.fastpeoplesearch.com
- https://www.truepeoplesearch.com
- https://www.familytreenow.com
- https://people.yandex.ru

People search websites allow to opt out, but after people remove themselves from listings, new search services appear with their records in them. The reason for that is the same dataset is bought and used by different services. Some companies own those datasets and even if on one of their websites a person removes the listing, on the new domain the old data is repopulated again so the previously removed profile reappears in the search. Consequently, if people did a pretty good at cleaning their stuff up you just have to wait for a new database to appear. One of the methods to find people that opted out is to go the people search service, find a unique paragraph, do a quoted Google search on it and find all of the domains that the company owns. There are chances that information your target removed from site A is now on site B.

. . .

User name

IntelTechniques.com OSINT Workflow Chart: User Name

Firstly, we have to find a username. Usually, it is a name plus surname combination or derived from the email, domain name of the website the person uses or owns. Start with data you have and do a reverse lookup towards what you need. Obviously, the simplest way is to Google any relevant data known to you at the moment and try to find any pages with the username. Also, you can use special websites that do a **reverse username search**, like socialcatfish.com, usersearch.org, or peekyou.com.

Google Dorks

The same Google Dorks that I showed for the real name search will be useful when searching for a username. In addition, URL search might give you good results as usually URLs contain usernames.

- *inurl:johndoe site:instagram.com*—search for URLs on Instagram that contain “johndoe” in them.
- *allinurl:john doe ny site:instagram.com* — find pages with “john”, “doe”, and “ny” words in the Instagram URL. Similar to *inurl* but supports multiple words.

Depending on the complexity of your search and how successful it was using previous methods you might want to generate a wordlist. It’s useful when you need to try a lot of options as you don’t have a clear picture of what username should be but have a lot of guesses. I have used this Python script for generating the wordlist below:

Name and surname were specified in Names.txt, in Terminal we just see the output

Username search

There are a lot of websites with a username search, I find these to be one of the best: instantusername.com and namechk.com. Usually, one service finds accounts that other one doesn't so it's better to use both websites. Apart from online services you can use WhatsMyName — a Github project, included in more advanced tools: Spiderfoot and Recon-ng. However, you can use it as a standalone checker running the Python script.

Searching for ”johndoe” username on 152 sites with WhatsMyName

While searching, you might get false positives as someone else can use the same username, be prepared for that.

Note: Running WhatsMyName, as well as any locally installed tool, could be an issue when you have certain websites blocked by the ISP. In that case, going through proxy or VPN will solve the issue. Moreover, to avoid exposure you should use anonymizers anyway.

· · ·

Email Address

IntelTechniques.com OSINT Workflow Chart: Email Address

Google Dorks

- *“@example.com” site:example.com* — search for all emails on a given domain.
- *HR “email” site:example.com filetype:csv | filetype:xls | filetype:xlsx* — find HR contact lists on a given domain.
- *site:example.com intext:@gmail.com filetype:xls* — extract email IDs from Google on a given domain.

Email tools

- Hunter — performs fast scan of the domain name for email addresses and reveals its common pattern.
- Email permutator — generates permutations of up to three domains at which target is likely to have an email address. Supports multiple variables input to generate custom results.
- Proofy — allows bulk email validation which is useful when you generated a list of emails using a permutation tool and want to check all of them at once.
- Verifalia — validates single email addresses for free without registration. To use bulk validation you have to sign up.

Browser plugins

- Prophet — reveals more information about people. It uses an advanced engine to predict the most likely email combination for a given person based on name, company and other social data. Then, Prophet verifies the generated email to make sure it is correct and deliverable.

- OSINT browser extension — contains a lot of useful links, including ones for email search and verification. Compatible with Firefox and Chrome.
- LinkedIn Sales Navigator — plugin for Chrome that shows associated Twitter account and rich LinkedIn profile data directly in Gmail.

Compromised databases

Data breaches have become a big issue and recently we are seeing more and more data dumps. Security researcher Troy Hunt collected released data, stripped off passwords, assigned emails to the breach they were involved in, and uploaded it to haveibeenpwned.com. While the fact of the breach itself might not be as important, what's important is with the email you might get a list of services that person uses or at least used.

Another option would be to use dehashed.com. With a free account it works similarly to Troy Hunt's website but with an active subscription it shows passwords in clear text or password hashes. From an OSINT perspective, we need that to search whether it was used on some other websites — one more way to find out which services the person uses or at least used. Doing the search by password or its hash shows not only on which website it was used, but also email address tied to it. Thus, we can get the target's emails we wouldn't obtain otherwise.

It’s important to note that if the password is not unique we might get false positives as other people might use it as well.

. . .

Phone number

IntelTechniques.com OSINT Workflow Chart: Telephone #

Sometimes people link a phone number and email to their Facebook profile, so typing it in the Facebook search might show you the profile. Another option is to look up user-supplied databases of phone numbers, like whocalledme.com. The database is not limited only to America, numbers from Europe can be checked as well. Besides, for those who want something like this but on the mobile device there are several apps: privacystar.com, getcontact.com, and everycaller.com. There are many reverse phone lookup services and they are usually country-specific so find the one that fits your need.

PhoneInfoga

PhoneInfoga is one of the most advanced tools to scan phone numbers using only free resources. The goal is to first gather basic information such as country, area, carrier, and line type on any international phone numbers with very good accuracy. Then try to determine the VoIP provider or search for footprints on search engines to try to identify the owner.

Features:

- Check if phone number exists and is possible
- Gather standard information such as country, line type, and carrier
- Check several numbers at once
- OSINT reconnaissance using external APIs, Google Hacking,

phone books, & search engines

- Use custom formatting for more effective OSINT reconnaissance

Well, you can see how many resources were scanned. Definitely faster than manual search.

Android Emulator

Many Android apps will work on an emulator without problems but some might not work as expected. For example, Viber has issues with VoIP phone numbers, as tested on freephonenum.com. However, there are many advantages running apps on the emulator: your real accounts or phone number will be safe as you don’t have to install questionable apps on your phone and you can easily spoof GPS coordinates.

Save the number in your phone and look at the Viber or WhatsApp contact list. These services allow adding a photo, biography, and name of the owner and this information can be extracted just by knowing the telephone number.

- Bluestacks — made primarily for gamers but runs other apps as well. Available for Windows, Mac and Linux and doesn't require a Virtual Machine to set it up so it installs easier than Genymotion.
- Genymotion — widely used by developers but also has a free version for personal use. Works on Windows, Mac and Linux and has a range of virtual devices to choose from. Use this guide from IntelTechniques to set up the emulator.
- AMIDuOS — available only for Windows and leverages device drivers from the system to enable near-native performance in Android. It's fast and has a straightforward installation. However, while the aforementioned emulators can be installed for free, AMIDuOS comes at a price of $10.

. . .

Domain name

IntelTechniques.com OSINT Workflow Chart: Domain Name

If the person or an organization owns a website you should know how to grab information about it. The investigation might reveal the operating system being used, software version, personal contact info, and more. I have to mention that it is advised to investigate without ever 'touching' the target's environment, such technique is called **passive reconnaissance** — footprinting that involves the uses of tools and resources that can assist in obtaining more information about your target without directly interacting with it. Below I describe methods of obtaining information while remaining hidden.

Google Dorks

Google Dorks is a passive information gathering method that was already mentioned above. Here I'm going to show what queries might be useful during domain investigation.

- *site:example.com* — limits search to a particular website or domain.
- *filetype:DOC* — returns DOC files or other specified types, such as PDF, XLS and INI. Multiple file types can be searched for simultaneously by separating extensions with "|".
- *intext:word1* — search for pages & websites that contain the specific word that you are searching.
- *allintext: word1 word2 word3* — search for all the given words in a page or website.

- *related:example.com* — will list web pages that are “similar” to a specified web page.
- *site:*.example.com* — show all subdomains. Asterisk acts as a substitute for a whole word or words in search queries.

Whois

Whois provides information about the registered users or assignees of an Internet resource, such as a domain name, an IP address block, or an autonomous system. There are many Whois resources, these are the good ones: whois.icann.org and whois.com.

Reverse Whois

Reverse Whois gives you a list of domains that have the same organization name or email address in their Whois record. For example, if you are investigating a company with the name “John Doe Inc.” you can see all the other domains registered under the “John Doe Inc.” One of my favorite websites is viewdns.info as it has an extensive toolkit, including reverse whois lookups.

Same IP

Often discovering what site is running on the same server as your target’s website uncovers valuable information. For example, you might find sub-domains or development sites. Often the service provider who hosts this site is responsible for other services as well — use www.atsameip.intercode.ca and www.sameip.org to check it.

Passive DNS

Using only DNS records you can see what IP resolved to the name or what name resolved to the IP. Sometimes that is not enough and that's where **passive DNS** records come in handy. They allow to check all the names that resolved to the researched IP, thus you can build a useful history of resolutions. My favorite product is RiskIQ Community Edition because it gives more information than just passive DNS. VirusTotal or SecurityTrails can be used for that purposes as well.

Internet archives and cache

The WaybackMachine can be used to find previous versions of webpages, enabling one to see how websites looked earlier or to recover deleted pages. Archive.today is another time capsule for web pages with the ability to manually add live url snapshots to the archive.

There are cases when deleted pages were not archived but are still cached by search engines. They can be found on cachedview.com or you can request the cached version with the following Google query: *cache:website.com*. Didn't find anything on Google? Check the cache of other search engines but keep in mind that the cache shows the last time the page was indexed. Therefore, you might get the page with missing images and outdated information.

You may also like visualping.io — a monitoring service that takes screenshots of the webpage at the selected time and sends you an email alert if something changes.

Reputation, malware, and referrals analysis

Reputation is important to know with whom you are dealing with and whether the website can be trusted. In case of any suspicions, a malicious activity check using free online tools might save you the trouble of opening the website in the VM or going through other precaution steps. Referral analysis is a search for inward and outward HTML links. Although doing the test on its own is not going to get you precise results, still, it's one of the methods that might show you connected domains.

- www.siteworthtraffic.com — analyses website traffic (users, page views) and estimates how much revenue it could generate through ads.
- www.alexa.com — analyses website traffic and competitors, shows what they are doing better and gives advice on SEO improvement.
- www.similarweb.com — analytics tool which provides deep information on website or mobile ranking, performance, the source of traffic, and more. On top of that, it does referral analysis.
- https://sitecheck.sucuri.net — scans websites for known malware, blacklisting status, website errors, and out-of-date

software.

- www.quttera.com — offers free malware scanning and provides a comprehensive report that includes malicious files, suspicious files, blacklisted status, and more.
- www.urlvoid.com — helps you detect potentially malicious websites. Also, it gives more information about the domain (IP address, DNS records, etc.) and cross-references it against known blacklists.

IoT search engines

IoT (Internet of Things) search engines show you devices connected to the cyberspace — think Google Search but for Internet-connected devices. Why is that useful? Instead of actively scanning ports and services with Nmap, for example, you can request already available information about open ports, applications, and protocols. Shodan.io is the most popular internet scanner with public API and integration with many security tools. For marketers, it provides data about product users and where they are located. Security researchers use it to uncover vulnerable systems and get access to a wide range of IoT devices. There are other alternatives like Censys, or its Chinese analogies — Fofa and ZoomEye.

. . .

Location search

IntelTechniques.com OSINT Workflow Chart: Location

Geolocation tools

Creepy is a free tool that allows data gathering from social networks and image hosting services for location research. The commercial option would be Echosec that costs nearly $500 per month.

IP-based Geolocation

IP-based Geolocation is a mapping of the IP address or MAC address to the real-world geographic location. There are many websites that map IP address to location, one of them is iplocation.net. When you know WI-FI access points the person has previously connected to — use wigle.net to map them and do more detailed research on Google Earth.

Useful websites

- www.emporis.com — database of architecture, provides images of buildings from all over the world. Might be useful to determine what building is on the picture.
- http://snradar.azurewebsites.net — search for geotagged public posts VKontakte and filter them by date.
- http://photo-map.ru — allows to search geotagged VKontakte posts, as a previous service, but requires authorization.
- www.earthcam.com — the global network of owned and operated live streaming webcams which might be useful during location research.
- www.insecam.org — a directory of online security cameras. The

coordinates of the cameras are approximate and point to the ISP address and not the physical address of the camera.

. . .

Images

When you have a picture and want to know where else it is used or when did it first appear — do a **reverse image search** using Google Images, Bing Images, and Baidu Images. In addition, TinEye's algorithms are designed differently than Google's and as such can return different results. Why is that useful? As an investigator, you may find the person by avatar, as people usually don't bother changing profile pictures for the various social networks they use. As a journalist, you may perform an image search paired with filtering to expose fake news. For example, a picture taken on the day of the event searched with date filter range that is earlier than the described event can't be found earlier. Thus, if the image is found — it was created before the event, therefore it's fake. If you need a narrow search across the social network, Findclone and Findmevk.com for Vkontakte or karmadecay for Reddit will do the job. Also, it's worth to mention browser extensions: RevEye for Chrome and Image Search Options for Firefox. Mobile apps like CamFind for iOS might help you searching for things from the physical world. Moreover, there is an Image Identification Project to identify what's on the image using AI.

The image itself contains a lot of useful information, like the camera information, geocordinates, etc. — it’s called **EXIF data** and if it wasn’t removed you might find a lot of interesting info. For example, with map geocordinates to find out where the picture was taken. With a camera serial number, you can look to see if there are pictures taken with that camera on the internet— stolencamerafinder.com. Image editing tools allow to view metadata. If you don’t want to install a complex program: Exiftool — the cross-platform free software might the thing you are looking for. The third option is to view EXIF data online: exifdata.com or viewexifdata.com. To remove EXIF data you can use a locally installed tool: exifpurge.com or do it online: verexif.com.

Do you need to perform **image forensics** and find out if the image was tampered with? Use Forensically or FotoForensics. If you don’t want to upload an image online — Phoenix or Ghiro can be run locally. The latter is more automated and gives you more functionality than the above mentioned online tools. Apart from that, working with images you might need to deblur it or improve the quality, so here are some enhancement tools:

- Smartdeblur — restores motion blur and removes Gaussian blur. Helps to restore focus and do image improvements which deliver amazing results.
- Blurity — focuses only on deblurring images, doesn't provide such many options like the previous tool and available only on Mac.

- Letsenhance.io — enhance and upscale images online using AI.

. . .

SOCMINT

SOCMINT is a subset of OSINT that concentrates on data gathering and monitoring on social media platforms. I have already described some social media intelligence techniques. Here I will complete the picture by listing more tools.

Facebook

- ExractFace — extracts data from Facebook, making it available offline to use as evidence or perform advanced offline analysis.
- Facebook Sleep Stats — estimates sleeping patterns based on users online/offline status.
- lookup-id.com — helps you to find the Facebook ID for a profile or a group.

Twitter

- Twitter advanced search — well, that's pretty self-explanatory :)
- TweetDeck — gives you a dashboard that displays separate columns of activity from your Twitter accounts. For example, you might see separate columns for your home feed, your notifications, your direct messages, and your activity — all in one

place on the screen.

- Trendsmap — shows you the most popular trends, hashtags, and keywords on Twitter from anywhere around the world.
- Foller — gives you rich insights about any public Twitter profile (profile public information, number of tweets and followers, topics, hashtags, mentions).
- Socialbearing — free Twitter analytics & search for tweets, timelines & twitter maps. Finds, filters, and sorts tweets or people by engagement, influence, location, sentiment, and more.
- Sleepingtime — shows the sleeping schedule of Twitter public accounts.
- Tinfoleak — shows devices, operating systems, applications and social networks used by the Twitter user. Also, it shows places and geolocation coordinates to generate a tracking map of locations visited. Maps user tweets in Google Earth and more.

Instagram

- www.picodash.com — exports followers statistics of a selected user or statistics by a selected hastag to a spreadsheet (CSV). Also, it exports likers and comments.
- https://web.stagram.com — online Image and Video Viewer/Downloader.
- https://codeofaninja.com/tools/find-instagram-user-id — gets

user ID. Usernames might change so it's useful to know profile's ID to not to lose the page.

- http://instadp.com — shows profile picture in full size.
- https://sometag.org — searches for trending hashtags, locations and accounts. In addition, it compares accounts and exports followers and hashtag statistics.

LinkedIn

- InSpy — an enumeration tool that is written in Python. Can be used to search for employees of a specific organization. Additionally, it can find out what technologies the organization uses, which is done by a crawling job listing for specific keywords.
- LinkedInt — scrapes e-mail addresses of employees in a selected organization. Supports automated e-mail prefix detection for a given company domain name.
- ScrapedIn — a Python script that scrapes profile data and imports it into XLSX file (intended to be used with Google Spreadsheets).

. . .

Automating OSINT

The Internet is an ocean of data and looking for the information manually might be time-consuming and not effective, plus automated tools could make correlations you wouldn't spot otherwise. It all depends on your case, whether you need to use these tools or not, as most of them have a steep learning curve and are required to solve complex problems. Thus, if you need to accomplish several simple tasks — don't bother installing software, just use online services and standalone scripts I have described earlier. To save some time and have an investigative environment ready, with all of these described below tools installed (excerpt FOCA), you can download Buscador OS — Linux Virtual Machine that is pre-configured for OSINT.

SpiderFoot

SpiderFoot is one of the best reconnaissance tools out there if you want to automate OSINT as it can be used to query more than 100 public data sources simultaneously and its modularity allows to fine-tune queried sources. What I personally liked is scanning by use cases. There are four different use cases: get everything and everything about the target, understand what your target exposes to the Internet (done through web crawling and search engine use), query blacklists and other sources to check target's maliciousness, and gather intelligence through different open sources. The last one is ideal for passive reconnaissance.

theHarvester

theHarvester is a very simple, yet effective tool used to fetch valuable information about the target on information gathering stage. It is great for scanning domain related information and harvesting emails. For passive reconnaissance, theHarvester uses many resources to fetch the data like Bing, Baidu, Yahoo, and Google search engines, and also social networks like LinkedIn, Twitter, and Google Plus. For active reconnaissance, it does DNS reverse lookup, DNS TDL expansion, and DNS brute force.

Recon-ng

Recon-ng is another great command line tool used to perform information gathering thoroughly and quickly. This full-featured Web Reconnaissance framework includes a good selection of modules for passive reconnaissance, convenience functions and interactive help to guide you on how to use it properly. For those familiar with Metasploit, Recon-ng will be easier to learn as it has a similar usage model. If you are looking for something powerful that can quickly check the visibility of your company on the Internet — this is the go-to tool.

People Recon With Recon-ng

Open source intelligence gathering guide

medium.com

Maltego

Maltego is an advanced platform developed for analyzing complex environments. Apart from data mining, it does data correlation and visually presents it. Maltego works with entities (people, companies, web sites, documents, and more) which you connect for further information gathering about them from different sources to get meaningful results. The distinctive feature of this tool is “transforms” — a library of plugins that help to run different kinds of tests and data integrations with external applications.

FOCA

FOCA (Fingerprinting Organizations with Collected Archives) is a tool for extracting hidden information and metadata from analyzed documents. When all documents are analyzed and metadata extracted it does automated metadata analysis to establish which documents were created by the same user. Also, it does correlation by server and printer. The latest version is available only for Windows.

Metagoofil

Metagoofil is a command line tool that is used to download public documents from websites with the following analysis and metadata extraction. It works with pdf, doc, xls, ppt, and other formats.

. . .

Conclusion

To conclude, it’s hard to stay private in the post-privacy world and control what information is floating in this digital ocean. While you can’t control everything that’s out there about you, it’s important to be at least aware about it. It goes without saying, that in the digital age, information plays a key role, so those who know how to find it will always be one step ahead. That’s what this article is for, to show how OSINT helps to solve a broad range of problems: from marketing to investigations to cybersecurity. However, I have described only the tip of the iceberg and most techniques in the article are simple but yet powerful. Therefore, some of the techniques when used in a malicious purpose might cause damage so I expect you will use them sensibly.

While this article was more about intelligence gathering, the next one will be about the preparation phase. Let me know in the comments if there is something specific you want to know about preparing an investigative environment. By the way, what tools and techniques do you use to gather intelligence?