2019 OSINT Guide

January 5, 2019 · 17 minutes read osint

I have been doing a lot of Open-Source Intelligence (OSINT) lately, so to celebrate 2019, I decided to summarize a lot of tips and tricks I have learned in this guide. Of course, it is not the perfect guide (no guide is), but I hope it will help beginners to learn, and experienced OSINT hackers to discover new tricks

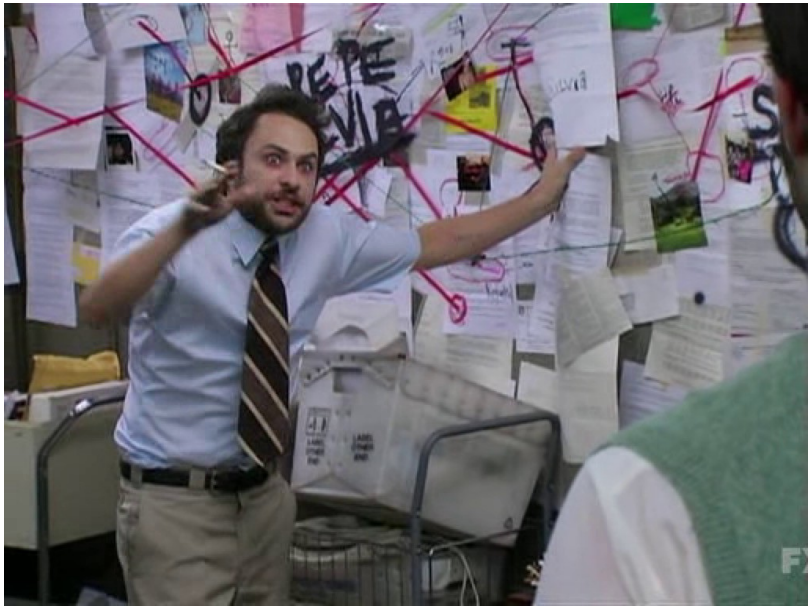
Methodology

The classic OSINT methodology you will find everywhere is strait-forward:

- Define requirements: What are you looking for?
- Retrieve data
- Analyze the information gathered
- Pivoting & Reporting: Either define new requirements by pivoting on data just gathered or end the investigation and write the report.

This methodology is pretty intuitive and may not help much, but I think it is still important to go back to it regularly, and take the time to make an interation of the loop. Very often during

investigations, we get lost into the amount of data gathered, and it is hard to have a view of what direction should the investigation take. In that case, I think it is helpful to take a break and go back to step 3 and 4: analyze and summarize what you have found, list what could help you pivoting and define new (or more precise) questions that still need answers.

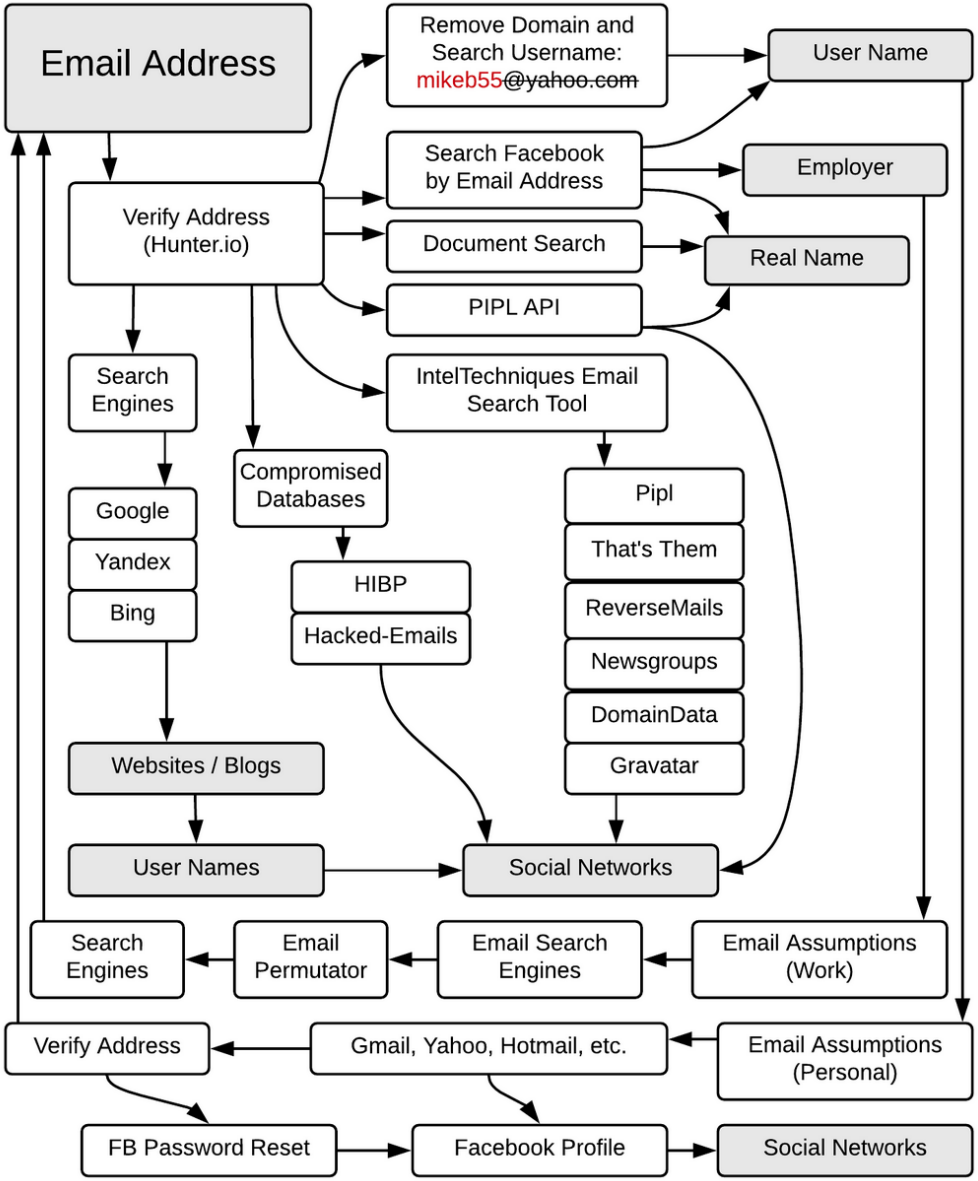


The other advices I would give are:

- **Never give up:** there will be a time where you have the feeling you have explored all the possibilities to get information. Don't give-up. Take a break (an hour, or a day doing something else), then analyze your data once again and try to see them with a different perspective. Is there a new piece of information you could pivot on? What if you asked the wrong questions at first? Justin Seitz recently wrote a blog post about tenacity giving a couple of examples where tenacity paid off.

- **Keep Evidences:** Information disappear online very quickly. Imagine you do a single opsec mistake, like clicking on a like on a tweet or the person you are reseaching start to be suspicious, suddently all the social media accounts and websites can disappear from one day to the other. So keep evidences: screenshots, archives, web archives (more information later) or anything else that works for you.
- **Timelines are good:** in forensic, timeline and pivoting on events happening in the same time is key. It is definitely not as important in OSINT but still a very interesting tool to organize your data. When was the website created? When was the FB account created? When was the last blog post done? Having all this in a table often give me a good view of what I am looking for.

Then there are two other methods I find useful. The first one are flowcharts to describe the workflow to search for more information based on a type of data (like an email). The best one I have seen are the one done by Michael Bazzell at IntelTechniques.com. For instance here is Michael Bazzell workflow when researching information on an email address:



Email OSINT WorkFlow by Michael Bazzell

After some time, I think it is a good idea to start developing your own investigation workflow and slowly improve it over time with new tricks you find.

The last methodology I would recommend for long investigations is the Analysis of Competing Hypotheses. This methodology was developed by the CIA in the 70's to help analyst remove bias from their analysis and carefully assess the different hypotheses. Bear in mind that it is a heavy and time-consuming tool, but if you are lost into a year long investigation, sometimes it is good to have a process helping you carefully evaluate your hypotheses.

Prepare Your System

Before jumping into the investigation, there are a couple of operational security aspects you should consider in order to avoid alerting the people you are researching about. Visiting an obscure personal website could give your IP address and hence your location to your target, using your personal social media account could lead to a click on a like by mistake. etc.

I follow the following rules when doing my investigations:

- **Use a commercial VPN or Tor** for all connections from your investigation browser. Most commercial VPNs provide servers in different countries and Tor allows you to choose the exit node country so I try to choose a country that would not raise a flag in that context (US for an investigation on a US organisation etc.).
- Do all the scans and crawling tasks from a cheap VPS that has *no link* with you.
- Use social media accounts dedicated to investigation and created under a fake name.

With all this done, you can now investigate as late in the night as you want, it is pretty unlikely that people will be able to identify who is looking for them.

Tooling

The question of tool is always a curious one in infosec, nothing bother me more than people listing endless list of tools in their CV and not skills they have. So let me say it clearly: **tools does not matter, it is what you do with tools that matter**. If you don't know what you are doing, tools won't help you, they will just give you a long list of data that you won't be able to understand or assess. Test tools, read their code, create your own tools etc, but be sure that you understand what they do.

The corollary of that is that there is not perfect toolkit. The best toolkit is the one you know, like and master. But let me tell you what I use and what other tools may be of interest to you.

Chrome and Plugins

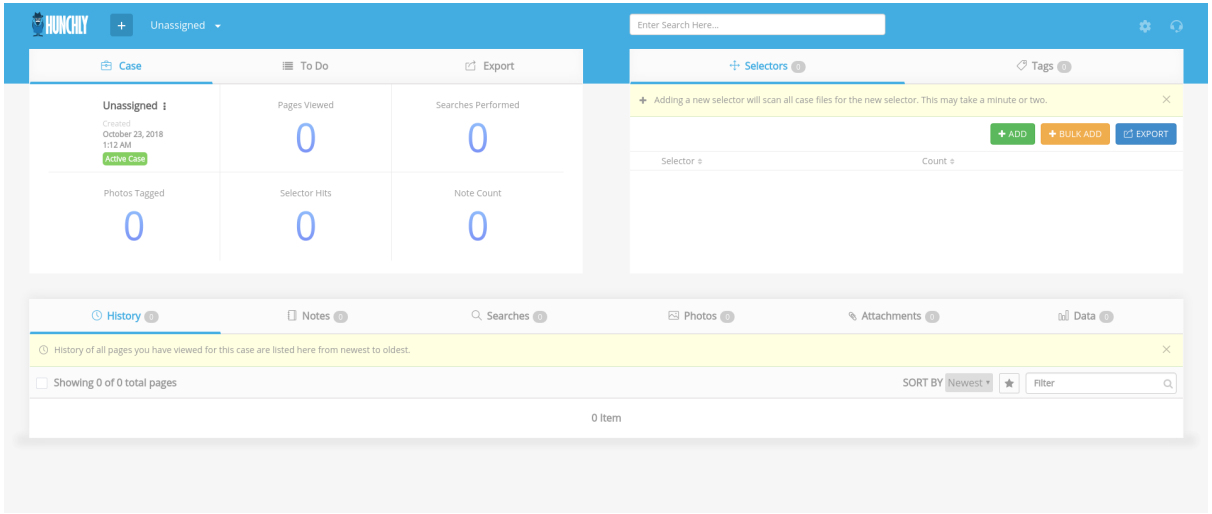
I use Chrome as my investigation browser, mostly because Hunchly is only available for Chrome (see after). I add to it some helpful plugins:

- archive.is Button allows to quickly save a webpage in archive.is (more about this later)
- Wayback Machine to search for archived page in the archive.org Wayback machine
- OpenSource Intelligence gives a quick access to many OSINT tools
- EXIF Viewer allows to quickly view EXIF data in images
- FireShot to take screenshot quickly

Hunchly

I recently started to use Hunchly and it is a great tool. Hunchly is a Chrome extensions that allows to save, tag and search all the web data you find during investigation. Basically, you just have to click on “Capture” in the extension when you start an investigation, and Hunchly will save all the webpages you visit in a database, allowing you to add notes and tags to them.

It costs USD130 / year, which is not that much considering how helpful it is.

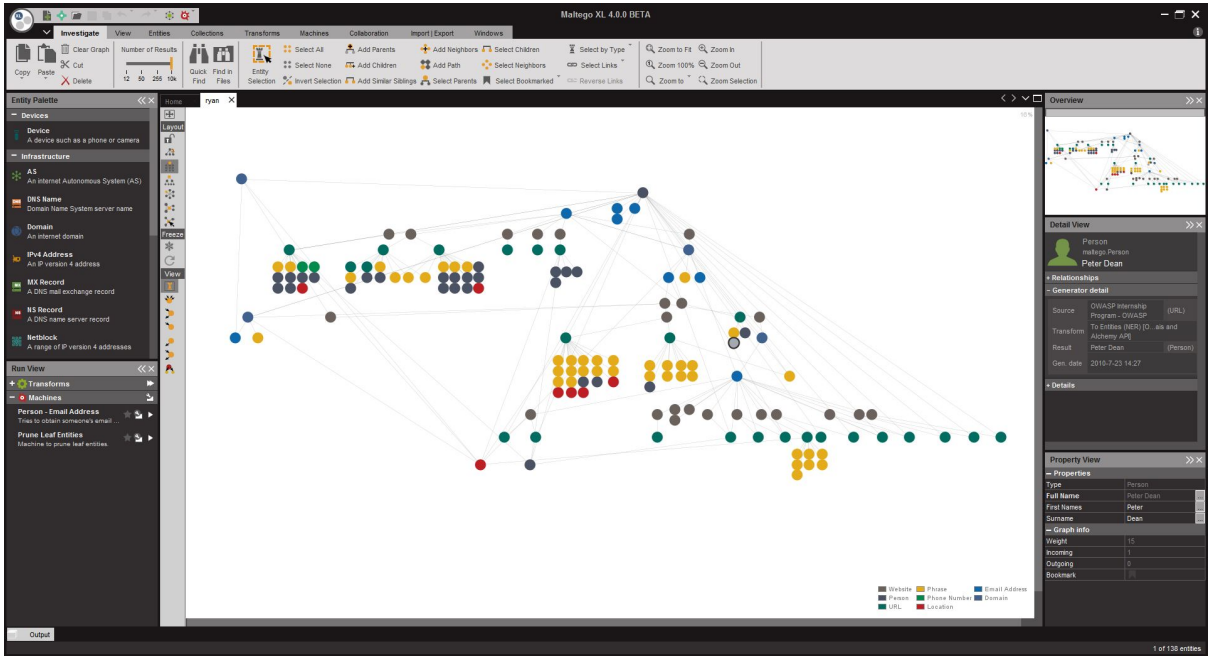


Screenshot of the Hunchly Dashboard

Maltego

Maltego is more a threat intelligence tool than an OSINT tool and has many limitations, but a graph is often the best way to represent and analyze investigation data and Maltego is good for that. Basically Maltego offer an GUI to represent graphs, and transforms to find new data in the graph

(for instance, domains linked to an IP address from a Passive DNS database). It is a bit expensive (USD999 / year the first year, then USD499 / year for renewal) and may only be worth it if you are also doing threat intelligence or a lot of infrastructure analysis. You can also use the Maltego Community Edition which limit the utilization of transform and the size of graph, but it should be largely enough for small investigations.



Screenshot of Maltego (source: Paterva)

Harpoon

I have developed a command-line tool called Harpoon (see the blog post here for more details). It started as a threat Intelligence tool but I have added many commands for OSINT. It is working with python3 on Linux (but MacOS and Windows should work too) and open source.

For instance, you can use Harpoon to search for a PGP key on key servers:

```
$ harpoon pgp search tek@randhome.io
[+] 0xDCB55433A1EA7CAB  2016-05-30      Tek__ tek@randhome.io
```

There is a long list of plugins, feel free to suggest or develop more or to create issues for new interesting features.

Python

Very often, you will end up with specific data gathering and visualization tasks that cannot be done easily with any tool. In that case, you will have to write your own code. I use python for that, any modern programming language would work equaly, but I like the flexibility of python and the huge number of libraries available.

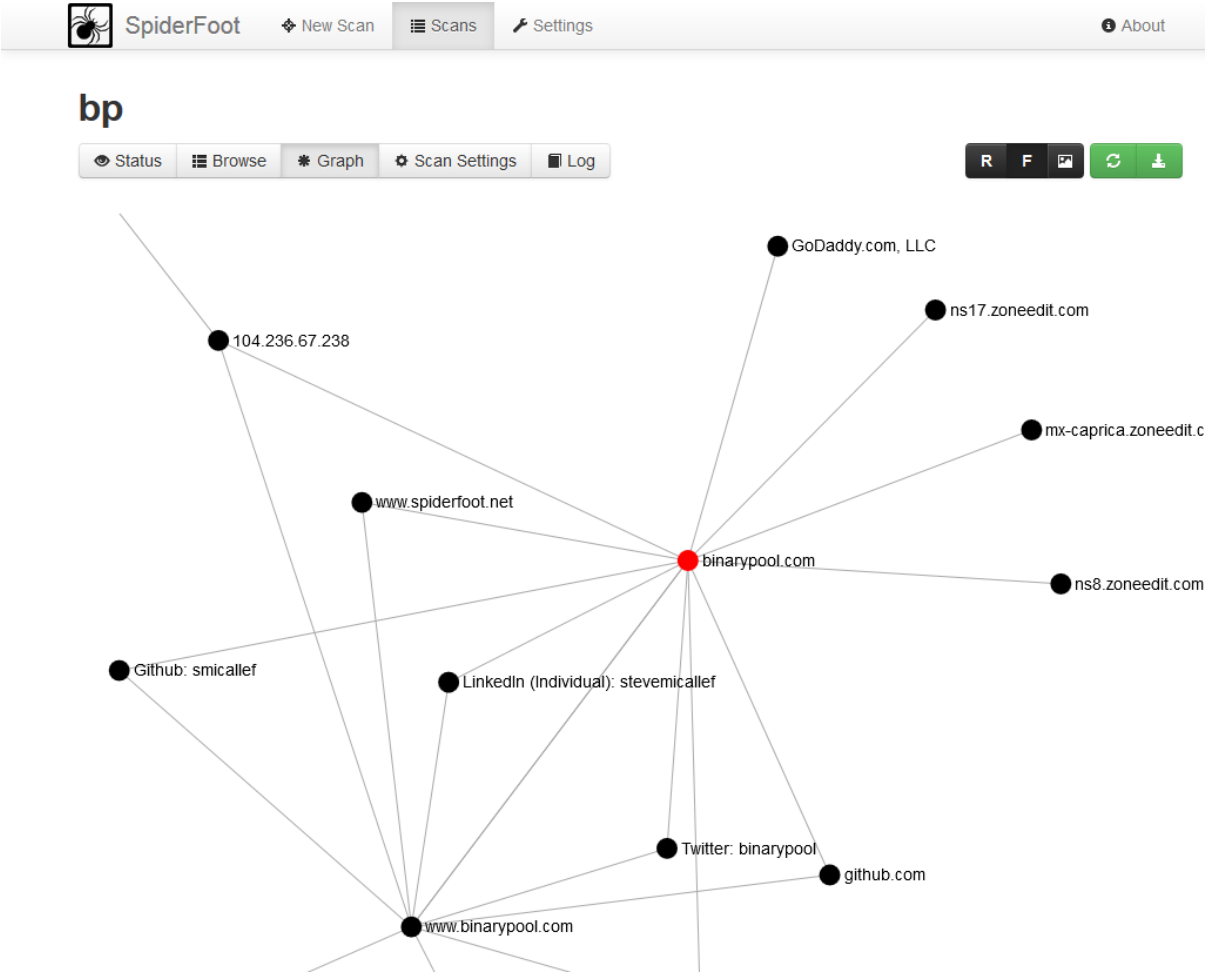
Justin Seitz (the author of Hunchly) is a reference on python and OSINT, and you should definitely have a look at his blog Automating OSINT and his book Black Hat Python

You may also like

There are many other tools for OSINT of course, but I find them less useful in my everyday work. Here are some tools you may want to check still, they are interesting and well done but do not really fit into my habits:

- SpiderFoot is a reconnaissance tool that gather information through many different modules. It has a nice web interface and generate graphs showing links between the different types of data. What I don't like about it is that it is thought as the magic tool finding everything for

you, but no tool will ever replace you to know what you are looking for and analyze the results. Yes, you will have to do the research by yourself and read the results one by one, and SpiderFoot does not help much on that. Good work and nice interface tho.



Screenshot of SpiderFoot (source: spiderfoot.net)

- recon-ng is a nice CLI tool to query different platforms, social media or threat intelligence platforms. It is pretty close to what Harpoon does in fact. I don't use it because I already use Harpoon that fits my need and I don't really like the shell interface it offers.
- Buscador is a Linux Virtual Machine that embeds lots of different OSINT tools. I always prefer having my own customized systems but it is a nice way to try new tools without the burden of installing them one by one.

Let's Go!

Let's now enter in the real topic: what can help you in OSINT investigations?

Technical Infrastructure

Analysis of the technical infrastructure is at the crossroad between threat intelligence and open source intelligence, but it is definitely an important part of investigations in some context.

Here is what you should look for:

- **IP and domains:** there are many different tools for that but I find Passive Total (now called RiskIQ) to be one of the best source of information. Free access gives you 15 query per day through the web interface and 15 through the API. I rely mostly on it but Robtex, HackerTarget and Security Trails are other good options.
- **Certificates:** Censys is a great tool, but the less known and less fancy crt.sh is also a very good certificate transparency database

- **Scans:** it is often useful to know what kind of services are running on an IP, you can do the scan yourself with nmap, but you can also rely on platforms doing regular scan of all IPv4 addresses for you. The two main platforms are Censys and Shodan, they both focus on different aspects (more IoT for Shodan, nmore TLS for Censys) so it is good to know and use both of them. BinaryEdge is a pretty new alternative to them but that is quickly evolving. More recently a similar Chinese platform called Fofa has been launched. Another source of information is Rapid7 Open Data but you will have to download the scan files and do research on your own. Finally, I find historical information on IP addresses to be a goldmine to understand the evolution of a platform, Censys only provide this data through paid plans (available for free for academic researchers) but Shodan provides this directly through the IP which is great ! Check the command `harpoon shodan ip -H IP` to see what it gives (you will have to pay Shodan for a life account).
- **Threat information:** even if not essential in OSINT, it is always interesting to check for malicious activities an a domain, IP or url. To do that, I mostly rely on Passive Total OSINT and projects and on AlienVault OTX
- **Subdomains:** there are many different ways to find a list of subdomains for a domain, from Google search (*site:DOMAIN*) to searching in alternate domains in certificates. PassiveTotal and BinaryEdge implement this feature directly, so you can jus query them to have a first list.
- **Google analytics and social media**: the last information that is really interesting, is to check if the same Google Analytics / Adsense id is used in several website. This technique was discovered in 2015 and well described here by Bellingcat. To look for these connections, I

mostly use Passive Total, SpyOnWeb and NerdyData (publicwww is another non-free alternative).

Search Engines

Depending on the context, you may want to use a different search engine during an investigation. I mostly rely on Google and Bing (for Europe or North America), Baidu (for Asia) and Yandex (for Russia and Eastern Europe).

Of course, the first investigation tool is search operators. You will find a complete list of these operators for Google here, here is an extract of the most interesting one:

- You can use the following boolean logical operators to combine queries: `AND`, `OR`, `+` and `-`
- `filetype:` allows to search for specific file extensions
- `site:` will filter on a specific website
- `intitle:` and `inurl:` will filter on the title or the url
- `link:`: find webpages having a link to a specific url (deprecated in 2017, but still partially work)

Some examples:

- `NAME + CV + filetype:pdf` can help you find someone CV
- `DOMAIN -site:DOMAIN` may help you find subdomains of a website
- `SENTENCE -site:ORIGINDOMAIN` may help you find website that plagiarized or copied an article

Additional readings:

- Mastering Google Search Operators in 67 Easy Steps
- Google Hacking Database

Images

For images, there are two things you want to know: how to find any additional information on an image and how to find similar images.

To find additional information, the first step is to look at **exif data**. Exif data are data embedded into an image when the image is created and it often contains interesting information on the creation date, the camera used, sometimes GPS data etc. To check it, I like using the command line ExifTool but the Exif Viewer extension (for Chrome and Firefox) is also really handy. Additionally, you can use this amazing **Photo Forensic website** that has many interesting features. (Other alternatives are exif.regex.info and Foto Forensics).

To find similar images, you can use either Google Images, Bing Images, Yandex Images or TinyEye. TinyEye has a useful API (see here how to use it) and Bing has a very useful feature letting you search for a specific part of an image. To get better results, it can be helpful to remove the background of the image, remove.bg is an interesting tool for that.

There is no easy way to analyse the content of an image and find its location for instance. You will have to look for specific items in the image that let you guess in which country it can be, and then do online research and compare with Satelite images. I would suggest to read some good investigations by Bellingcat to learn more about it, like this one or this one.

Additional readings:

- Metadata: MetaUseful & MetaCreepy by Bellingcat
- The Visual Verification Guide by First Draft news

Social Networks

For social network, there are many tools available, but they are strongly platform dependent. Here is a short extract of interesting tools and tricks:

- **Twitter:** the API gives you the exact creation time and tool used to publish tweets. x0rz' tweets_analyzer is a great way to have an overview of the activity of an account. There are ways to find a Twitter id from an email address but they are a bit tricky.
- **Facebook:** the best resource for Facebook investigation is Michael Bazzell's website, especially his custom FB tool's page
- **LinkedIn:** the most useful trick I have found in LinkedIn is how to find a LinkedIn profile based on an email address.

Cache Platforms

There are several platforms caching websites that can be a great source of information during an investigation, either because a website is down or to analyse historical evolution of the website. These platforms are either caching websites automatically or caching the website on demand.

Search Engines: most search engines are caching websites content when they crawl them. It is really useful and many websites are available that way but keep in mind that you cannot control

when it was cached last (very often less than a week ago) and it will likely be deleted soon, so if you find anything interesting there, think about saving the cached page quickly. I use the following search engines cache in my investigations: Google, Yandex and Bing

Internet Archive: the Internet Archive is a great project that aims at saving everything published on Internet, which include crawling automatically webpages and saving their evolution into a huge database. They offer a web portal called Internet Archive Wayback Machine, which is an amazing resource to analyze evolutions of a website. One important thing to know is that the Internet Archive is removing content on demand (they did it for instance for the Stalkerware company Flexispy), so you have to save content that need to be archived somewhere else.

Other manual caching platforms: I really like archive.today that allows to save snapshot of webpages and look for snapshots done by other people. I mostly rely on it in my investigations. perma.cc is good but offer only 10 links per months for free accounts, the platform is mostly dedicated to libraries and universities. Their code is open-source tho, so if I had to host my own caching platform, I would definitely consider using this software.

Sometimes it is annoying to manually query all these platforms to check if a webpage was cached or not. I have implemeted a simple command in Harpoon to do that:

```
§ harpoon cache https://citizenlab.ca/2016/11/parliament-keyboy/
Google: FOUND https://webcache.googleusercontent.com/search?q=cache%3Ahttps%3A%2F%2Fcitizenl
ab.ca%2F2016%2F11%2Fparliament-keyboy%2F&num=1&strip=0&vwsrc=1
Yandex: NOT FOUND
Archive.is: TIME OUT
Archive.org: FOUND
-2018-12-02 14:07:26: http://web.archive.org/web/20181202140726/https://citizenlab.ca/2016/1
```

```
1/parliament-keyboy/
Bing: NOT FOUND
```

Also, keep in mind that Hunchly mentioned earlier automatically save a local archive of any page you visit when recording is enabled.

Capturing Evidences

Which bring us to the next point: capturing evidences. Capturing evidences is a key part of any investigation, especially if it is likely gonna be a long one. You will definitely be lost into the amount of data you found several times, web pages will change, Twitter accounts will disappear etc.

Things to keep in mind:

- You cannot rely on the Internet Archive, use other cache platforms and if possible on local copies

- Save images, documents etc.
- Take screenshots
- Save data about social media accounts, they can be deleted anytime. (For Twitter accounts, Harpoon has a command to save the tweets and user infos in a JSON file)

Additional readings:

- How to Archive Open Source Materials by Aric Toler from Bellingcat

URL shorteners

URl shorteners can provide very interesting information when used, here is a summary on how to find stats information for different providers:

- **bit.ly:** add a `+` at the end of the url, like https://bitly.com/aa+
- **goo.gl:** (soon deprecated), add a `+` at the tend which redirects you to a url like `https://goo.gl/#analytics/goo.gl/[ID HERE]/all_time`
- **ow.ly** is the url shortener of hootsuite but you cannot see stats
- **tinyurl.com** does not show stats but you can see the url with `http://preview.tinyurl.com/[id]`
- With **tiny.cc**, you can see stats by adding a `~`, like `https://tiny.cc/06gkny~`
- With **bit.do** you can add a - in the end, like `http://bit.do/dSytb-` (stats may be private)
- **adf.ly** is offering to make money by showing ads when redirecting to the link. They use a lot of other subdomains like `j.gs` or `q.gs` and do not show public stats
- **tickurl.com:** Access stats with + like `https://tickurl.com/i9zkh+`

Some url shorteners are using incremental ids, in that case it is possible enumerate them in order to find similar urls created around the same time. Check this good report to see an example of this idea.

Company Information

Several databases are available to search for information on a company. The main one is Open Corporates and OCCRP database of leaks and public records. Then you will have to rely on per country databases, in France societe.com is a good one, in the US you should check EDGAR and in UK, Company House (more information on it here).

Resources

Here are some interesting resources to learn more about OSINT:

- Bellingcat amazing resources: the guides and the community built list of resources
- The Github repository Awesome OSINT
- The OSINT framework
- osint.link

That's all folks, thanks for taking the time to read this post. Please feel free to help me complete that list by contacting me on Twitter.

This blog post was written mainly while listening to Nils Frahm.

Update 1: Add Yandex Images, remove Background and Photo Forensic. Thanks to Jean-Marc Manach and fo0 for their tips.