487.2

Gathering, Searching, and Analyzing OSINT

PLEASE READ THE TERMS AND CONDITIONS OF THIS COURSEWARE LICENSE AGREEMENT ("CLA") CAREFULLY BEFORE USING ANY OF THE COURSEWARE ASSOCIATED WITH THE SANS COURSE. THIS IS A LEGAL AND ENFORCEABLE CONTRACT BETWEEN YOU (THE “USER”) AND SANS INSTITUTE FOR THE COURSEWARE. YOU AGREE THAT THIS AGREEMENT IS ENFORCEABLE LIKE ANY WRITTEN NEGOTIATED AGREEMENT SIGNED BY YOU.

With the CLA, SANS Institute hereby grants User a personal, non-exclusive license to use the Courseware subject to the terms of this agreement. Courseware includes all printed materials, including course books and lab workbooks, as well as any digital or other media, virtual machines, and/or data sets distributed by SANS Institute to User for use in the SANS class associated with the Courseware. User agrees that the CLA is the complete and exclusive statement of agreement between SANS Institute and you and that this CLA supersedes any oral or written proposal, agreement or other communication relating to the subject matter of this CLA.

BY ACCEPTING THIS COURSEWARE, YOU AGREE TO BE BOUND BY THE TERMS OF THIS CLA. BY ACCEPTING THIS SOFTWARE, YOU AGREE THAT ANY BREACH OF THE TERMS OF THIS CLA MAY CAUSE IRREPARABLE HARM AND SIGNIFICANT INJURY TO SANS INSTITUTE, AND THAT SANS INSTITUTE MAY ENFORCE THESE PROVISIONS BY INJUNCTION (WITHOUT THE NECESSITY OF POSTING BOND) SPECIFIC PERFORMANCE, OR OTHER EQUITABLE RELIEF.

If you do not agree, you may return the Courseware to SANS Institute for a full refund, if applicable.

User may not copy, reproduce, re-publish, distribute, display, modify or create derivative works based upon all or any portion of the Courseware, in any medium whether printed, electronic or otherwise, for any purpose, without the express prior written consent of SANS Institute. Additionally, User may not sell, rent, lease, trade, or otherwise transfer the Courseware in any way, shape, or form without the express written consent of SANS Institute.

If any provision of this CLA is declared unenforceable in any jurisdiction, then such provision shall be deemed to be severable from this CLA and shall not affect the remainder thereof. An amendment or addendum to this CLA may accompany this Courseware.

SANS acknowledges that any and all software and/or tools, graphics, images, tables, charts or graphs presented in this Courseware are the sole property of their respective trademark/registered/copyright owners, including:

AirDrop, AirPort, AirPort Time Capsule, Apple, Apple Remote Desktop, Apple TV, App Nap, Back to My Mac, Boot Camp, Cocoa, FaceTime, FileVault, Finder, FireWire, FireWire logo, iCal, iChat, iLife, iMac, iMessage, iPad, iPad Air, iPad Mini, iPhone, iPhoto, iPod, iPod classic, iPod shuffle, iPod nano, iPod touch, iTunes, iTunes logo, iWork, Keychain, Keynote, Mac, Mac Logo, MacBook, MacBook Air, MacBook Pro, Macintosh, Mac OS, Mac Pro, Numbers, OS X, Pages, Passbook, Retina, Safari, Siri, Spaces, Spotlight, There’s an app for that, Time Capsule, Time Machine, Touch ID, Xcode, Xserve, App Store, and iCloud are registered trademarks of Apple Inc.

PMP and PMBOK are registered marks of PMI.

SOF-ELK® is a registered trademark of Lewes Technology Consulting, LLC. Used with permission.

SIFT® is a registered trademark of Harbingers, LLC. Used with permission.

Governing Law: This Agreement shall be governed by the laws of the State of Maryland, USA.

SEC487.2 | Open Source Intelligence (OSINT) Gathering and Analysis

SANS

Gathering, Searching, and Analyzing OSINT

Welcome to SEC487 Open Source Intelligence (OSINT) Gathering and Analysis!

487.2 Table of Contents

This table is a reference for you to quickly move to certain topics in this 487.2 book.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. **Data Analysis Challenges**
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

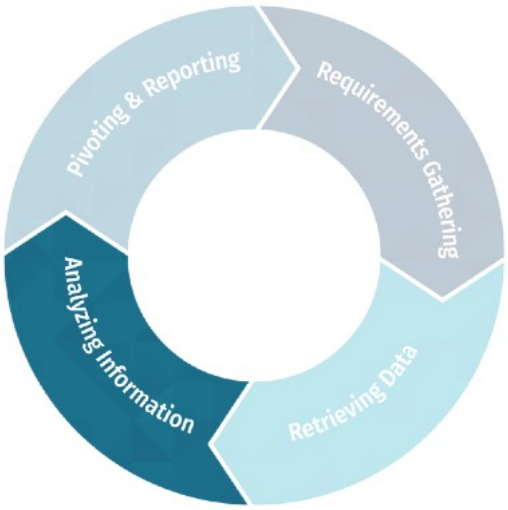
This page intentionally left blank.

Data Analysis

We are now going to discuss the 3rd section of the OSINT process

Going through all the data and evaluating it

Then deciding next steps



Data Analysis

On day 1 of the course, we learned of the sections of the OSINT process shown in the slide above. We need to dig deeper into the third piece of the process: data analysis. This crucial step is frequently skipped in rushed investigations and can lead to dangerous consequences, as we will see later.

After retrieval of data, we must analyze it to find patterns, understand what it means within the context of our assessment, and figure out other aspects of the target that we wish to investigate. Depending on the amount of data that was collected, parsing through it to find patterns and meaning can be challenging in some cases and frustrating in others. With API keys allowing us to pull hundreds or thousands of records about our target, data analysis may take some time to do thoroughly. Successful analysts understand this and plan ample time for data review. They record observations and then choose new areas to explore.

Data Analysis Process

When you evaluate data, ask yourself questions about the data collected:

- Is it most likely true?
- Could it be false?
- Is it helpful and relevant to the assessment?
- Do I have corroborating data from other trusted sources?

Even trusted sources can have "bad" data in them

Data Analysis Process
During the evaluation of the data you collected, it is good practice to come up with questions to ask yourself to ensure that your analysis is productive. Think about the social media sites you might use. Do any of them check to ensure you actually visited the places you "check into"? What about the place you work or schools you attended? Are any of those validated? Probably not. Even trusted repositories of data can have accidentally inaccurate or just wrong data in them. Treat all data as suspect until confirmed.

Some sample questions to start out the process are listed below. Carefully consider each as you scrutinize the information about your target:

- **Is it most likely true?** Data on the internet does not necessarily need to be true to be posted. Evaluate if what you found is more likely to be true than to be false. Some organizations have a rating scale that shows the truthfulness of the data or the confidence they have that the data is true. You need to evaluate the data with the same objective as the political fact-checking organization PolitiFact (https://sec487.info/41). They have a "truth-o-meter" with several categories of truthfulness that range from "true" to "pants on fire" (not true).
- **Could it be false?** This is the converse of the above question and is good practice. Instead of asking "is it true?" we look at the data and ask if it could be false. Examining information from different perspectives can help shed light on trends and help sort useful from less-useful information.
- **Is it helpful and relevant?** Many times we, as analysts, don't know where our investigations will take us. We gather and record all the data we can and then, when we get to this point in our process, we need to ask if this data is useful to the overall investigation. If it is, continue to evaluate it. If it isn't, move to other tasks.
- **Is there corroborating data?** When multiple sources state the same information, it can seem like the data should be more trustworthy, and depending on the sources of that data, this can be the case. But sometimes multiple sources get and post false data, which can lead investigators to draw incorrect conclusions.

Image from https://sec487.info/k2, March 25, 2018.

Inaccurate Data

Potential reasons for inaccurate data:

- Active measures by your target (Planting a false trail)
- Unintentionally incorrect data (OCR, Siri, typos)
- Old data that wasn't updated (whois)
- Similar target's data (Multiple people, same names)
- Human nature (Bragging, appearing to be someone they are not – think online dating profiles)

Inaccurate Data

While performing your OSINT assessments, you will gather data that is seemingly contradictory and maybe just plain wrong about your targets. You will suspect it is false due to various factors, such as inconsistency with existing, more-reliable data, or perhaps it is impossible for your target to have posted or experienced what you found. There are several other explanations for why you find false data:

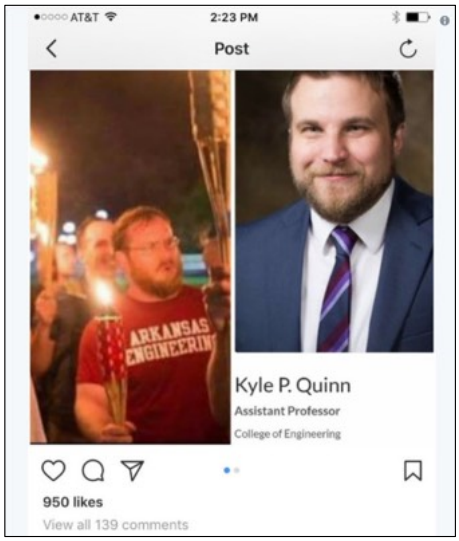
- **Active measures by your target** - Here we have a target that has intentionally placed false data where people and tools can find it to make it more challenging for people to understand the true state or nature of the subject. This may include setting up false social media accounts with the same name as the target as well as creating a variety of email addresses that could be plausible for the target to own and use. Depending upon the level of effort from your target, a campaign of this type can be extremely effective at confusing the real data with the false data.
- **Unintentionally incorrect data** - Humans enter data into computers and fill in forms. Our accuracy is not always 100% when typing and transcribing content from the paper world to the digital. Additionally, sometimes we use OCR (Object Character Recognition) audio assistants (e.g., Siri, Alexa) and other alternative forms of data entry that can produce data that is not accurate. While not intentional, these mistakes can make OSINT information analysis challenging.
- **Old data that was not updated** - The whois system is an excellent example of this class of false data. At one point, when a certain domain was registered, a person's or company's data was submitted to a domain registrar. As the years go by, the domain is consistently renewed by the person or company but maybe that registrant's email address or phone changes. There is no requirement to update the whois system with the accurate content and so that data was correct but no longer is.
- **Similar target's data** - This is a common occurrence for many of our assessments. Your target's name is very common for the region you are searching. When you search for your target, you receive plausible data, but it may be from other people. Here we need to perform additional analysis on the content to match details with known data.
- **Human nature** - In general, people want to present themselves in the "best" light and, therefore, may not wish to post to social media negative things about themselves. Think about what a person might put into an online dating profile that may "stretch" the truth a little bit.

Jumping to Conclusions

Here we gather some data and then reach an end state without substantial facts and/or using flawed logic

Example:

- Charlottesville, Virginia white supremacist rally in August 2017
- Caucasian male with beard marches with supremacists
- Innocent man identified and doxxed



Jumping to Conclusions

With this data analysis issue, someone has gathered some data. In their analysis of the content, they either do not have enough data to analyze or they apply flawed logic to the analysis and reach the wrong conclusions. It is usually when someone acts upon these incorrect conclusions that the real-world problems begin.

An excellent (and terrible) example of this occurred in August 2017 during the White Supremacist and Nazi rally in Charlottesville, Virginia. As reported by Business Insider,[1] people on the internet were trying to identify the participants in the rally through images from the videos shot at the event. Unfortunately, an innocent person, Kyle P. Quinn, was falsely accused of participating in the rally because he looked like a person on screen and he was an assistant professor at the university shown on the white supremacist's shirt. The article mentions that "after being wrongly identified, Quinn received countless threats online, with users calling him a racist and posting his home address and personal information online."

Jumping to conclusions in your investigations can have dire impacts. Be aware of it and work all hypotheses.

Reference and image:

[1] https://sec487.info/8b, August 16, 2017.

Bias

Personal prejudices affect analysis

- Are you researching something that goes against your beliefs?
- Judging your target because they violate your taboos/ethics?

In choosing sources to search/examine

- Seek to use standard, objective sources

In choosing what gets recorded/stored during collection for analysis

Confirmation Bias

Bias

Each of us has our own opinions, morals, and ethics. We gather these perceptions and moral compasses throughout our lives based upon our cultures, our education, our families, and our experiences. These things make you who you are. They also can taint your analysis of data.

Think about how your analysis might change if you hold a strong personal belief that, for example, pornography is morally wrong and degrading to certain people. Does this mean that you will not search for your OSINT target profiles on pornography web sites? Will you judge the target if you see nude pictures of them in their social media accounts? It can happen.

What if we hold a certain belief about our target and only seek sources and gather data to confirm that belief or to confirm our bias? These biases may cause us to not visit certain sites or use certain applications during our assessments, and this could cause us to miss valuable data about the target. Without that data, our analyses are inaccurate, and our final intelligence might be wrong.

Seek to use a variety of standard sources. Rely on your standard operating procedures for what data you need to gather and report from where. If you get to a place where you are uncomfortable with where your investigation is taking you or the data you are gathering, consult your colleagues and management to get direction.

False Generalization

Committing errors in drawing conclusions on the basis of data that is not representative of the population

Examples:

- "People from that country use that website"
- "People of that race are of that socio-economic level"
- "People in that country use smartphones"
- "Everyone has fast internet"

False Generalization

As OSINT analysts, we need to understand that while, in general, people from certain regions may use certain applications, just because someone is from one of those regions does not mean that they MUST use one of those sites. We need to move past the idea that everyone does something. For instance, in most countries people pay government taxes. We also know that some people do not, even though it is illegal. Making assumptions about classes of people will force your investigations to possibly avoid certain areas of relevant information because of your assumptions. This data analysis challenge is related to "Jumping to Conclusions."

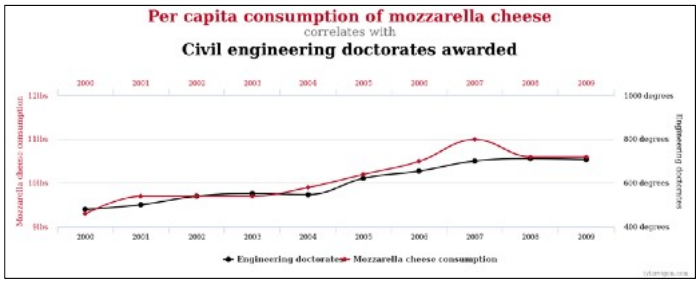
Stay open to the facts. Investigate all the angles and aspects without assuming—or if you do assume, cover your bases and ensure that you check out competing hypotheses as well.

Correlation Does Not Imply Causality

Correlation - There is a relationship between two or more events or items—either positive or negative

Causality - One or more things causes/makes something happen. Direct relationship. Cause and effect.

If correlation, then <u>maybe</u> causation



Example: Ashley Madison data dump. User account email addresses did not mean people were cheating.

Correlation Does Not Imply Causality

Correlation means that a relationship exists between two or more things. It could be a positive relationship where, as the number/amount/rate of one thing increases, the number/amount/rate of the other does the same. Or it can be a negative correlation where, as one gets bigger, the other gets smaller. Relationships are tricky and do not imply causation. Just because one thing goes up and another thing rises at the same rate, it doesn't mean that one causes the other's behavior. For an excellent example of this, check out the graphic above,[1] which shows a high correlation between things that do not cause each other (that is, the per capita consumption of mozzarella cheese and the numbers of civil engineering doctorate degrees awarded). There may be causation with correlation, but this does not have to be the case.

Causality is different. With causality, one thing makes something else happen. Many times, people assume that because two or more things may be highly correlated, they caused each other. Take, for example, user account email addresses found in the data dump from the Ashley Madison web site hack in 2015 (https://sec487.info/8u). The site helped people find others who wanted to have extra-marital affairs. User accounts/email addresses were dumped to the internet and pointed to as "people that are cheaters." What people didn't consider was that, when signing up for an account on Ashley Madison, you didn’t have to provide a real email address. They didn't check. So, you could have signed up with another person's email address. When the data went public, if I had used your email address in my Ashley Madison account, people would falsely assume YOU cheated on your spouse.

Reference: https://sec487.info/8t, October 7, 2017.

False Dichotomy

"Arbitrarily reducing a set of many possibilities to only two"[1]

We see this when people have challenges creating other causes or options

They reduce the possibilities to "this or that"

Sometimes referred to people seeing in "black and white"

Be open to and challenge yourself to discover alternatives to binary thinking

False Dichotomy

With a false dichotomy, we have something that has occurred and the person analyzing it creates two competing theories. They do not accept that alternative explanations for the events could be possible and, instead, only focus on the "this" or the "that." This is also referred to as binary thinking, since it is either “on” or “off” (a 1 or a 0).

We need to challenge ourselves to create alternative hypothesizes and explore them so that our analyses can be as complete and correct as possible.

Reference:

[1] https://sec487.info/8v

Discarding Unfavorable Data

Some of the data you acquire may not seem to "fit" into your explanations of events or about your target

Some people will discard that seemingly fraudulent or irrelevant data and not report it

Just because it is different, if it is trusted, it can be relevant with further data or analysis

Discarding Unfavorable Data

When performing an OSINT investigation, we will come across data that is from a trustworthy source but does not fit with our subject's profile or with the events that we have already pieced together. Some people will ignore or not report that data to their customers because it doesn't fit with their understanding.

We should be aware of this and remember to report all relevant data, even if it doesn't seem to fit right now. With additional data or with further analysis, we may find that the bit of information does fit.

Reference: https://sec487.info/8w

Sound Analysis

- Work with the facts and what we have found
- Understand the confidence of the sites we used
- Use corroborating data to bolster confidence
- Search for false data and disinformation
- Avoid the fallacies

Recognize that the facts may or may not be what your customer wants to hear

Sound Analysis

How do we ensure that the intelligence we provide to our customers is accurate, valid, and trustworthy? Using the facts that you gathered, you will perform analyses on them and create conclusions or hypotheses. Rely on and more heavily weigh the facts from trusted sources over those of suspicious integrity. Use corroborating data to increase your confidence that the data is accurate, and ensure that the additional data you are using is pulled from a different source that is not relying on the same data repository as the original content.

Be aware that the data you gather may be false, and work from the premise of confirming all important content prior to using it. Recognize also that your cognitive biases and mental fallacies can play into your analyses and skew outputs to your customers. Work to recognize and avoid (or at least work through) these issues.

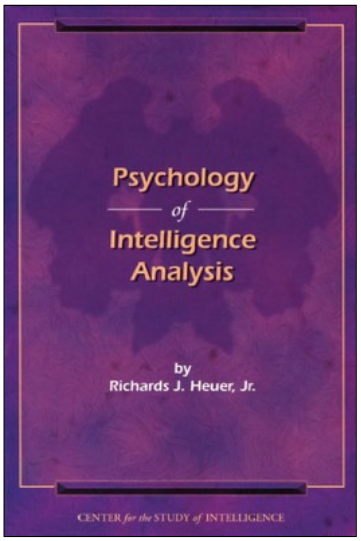
Long-time OSINT analyst and SEC487 instructor, John TerBush noted in a private conversation to the author, "a true investigation is conducted to discover the facts; these may or may not be what you or your client 'want' but if you wish to conduct fact-based investigations you should collect information without bias or self-selection. Sound analysis similarly should not be conducted 'from an angle.'"

Free PDF on Analytical Thinking

In 2007, the CIA released a PDF of Richards J. Heuer, Jr.'s work

The *Psychology of Intelligence Analysis* provides insights into how we analyze

It dives into bias and promotes methods like Analysis of Competing Hypotheses (ACH) to reach sound analysis



Free PDF on Analytical Thinking

A free PDF resource on thinking analytically comes from the CIA (https://sec487.info/e3). While it focuses on intelligence and not specifically OSINT, the underlying content on how to analyze data is similar. The Table of Contents is shown below.

Table of Contents

What If There Is No Data?

- What if you do all the right things and you find nothing?
- Does it "mean" something?
- Is it odd that the target has little to no internet footprint?
- What do you report?
- How long do you keep going?

What If There Is No Data?

You will encounter situations where the person or entity you are researching has a very small or no internet footprint. All of your searches and database queries show no results for the target. In these cases, you need to think about what this "means."

Is the target a fake persona? Are they someone who lives "off the grid" and owns no credit cards, has no social media accounts, and doesn't use smartphones or computers? Is it odd that you cannot find data?

There are other questions that you and your team need to think about too. Things such as, how long do you keep searching for the target when you are not receiving any valid results? If you find no data about your target, what do you report to your customer?

Being aware of the absence of data is important, especially if the target lives in a place where you would expect to see some internet data about them. What usually occurs in these cases is that we move from solid, trusted sources to more-obscure, perhaps less-trusted ones. If we don't find anything at all, that itself is a result to report to our customer.

Image from https://sec487.info/k3.

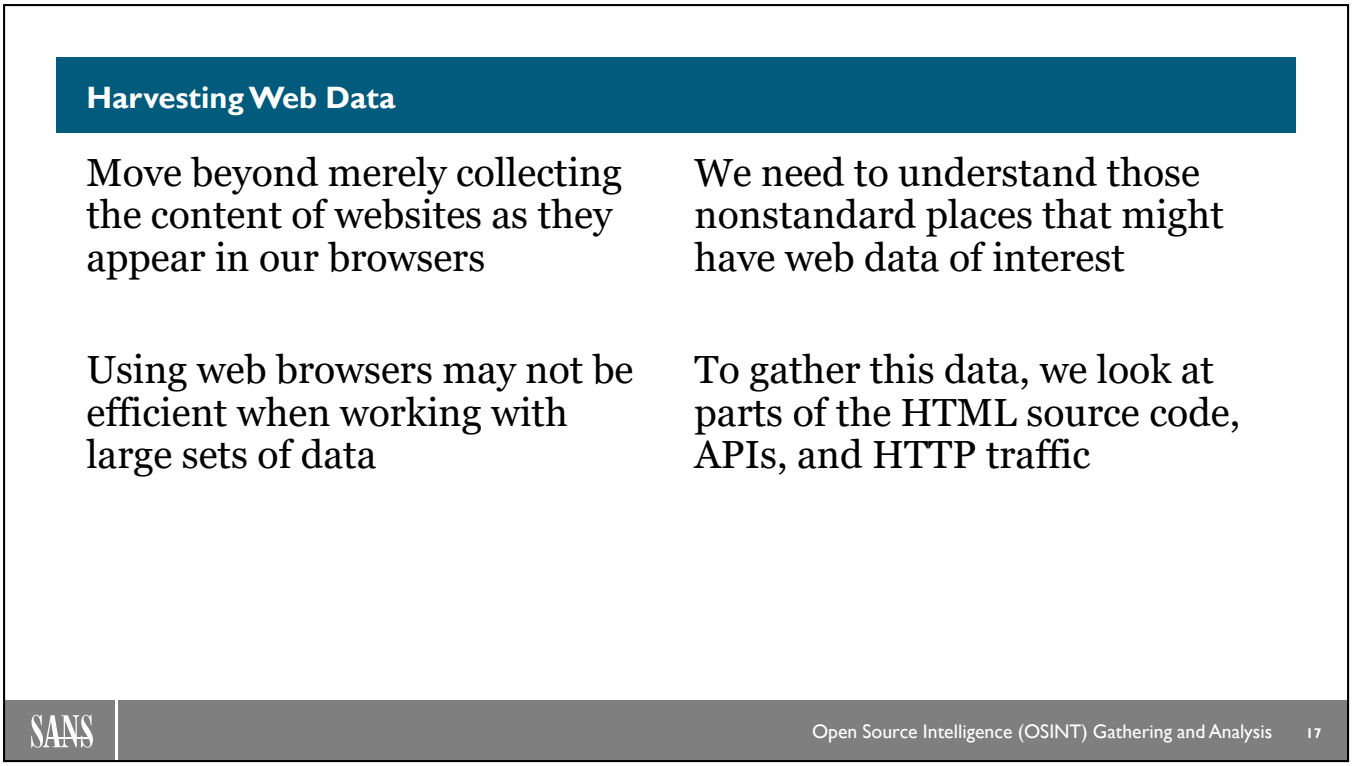
Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. **Harvesting Web Data**
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.



Harvesting Web Data

Much of our investigations take place within the confines of a web browser as we travel around the internet and dark web looking for the information our customers ask us to obtain. But only using a web browser to collect this data may not be efficient when we need to work with tens, hundreds, or thousands of records. Additionally, malicious content such as JavaScript cryptocurrency miners and browser plugin exploits can be used against our analyst systems. We need to keep our systems secure. It is important to understand and, when appropriate, employ a variety of techniques in retrieving the OSINT data we need.

Once we have the tools to gather the data we need, we must be able to find those places where the data resides. OSINT data may not only be displayed in the content of the pages that render in our browsers (and now, other tools), but it can also be found in the HTML source code of the pages and in other places on the web server.

When we put these two concepts together—understanding how to gather data from the internet and where that data is—we can better serve our clients by going deeper into less-common sources and methods.

The Retrievable Web

Much of the data passing between your browser or mobile device and a web site is text

There are "expected" methods of retrieving the data (browsers)

Sometimes other methods are more beneficial and safer

Additional methods of retrieving data:

- Using a proxy web application
- Using command-line tools
- Using a script
- Using an API
- Using cached content

The Retrievable Web

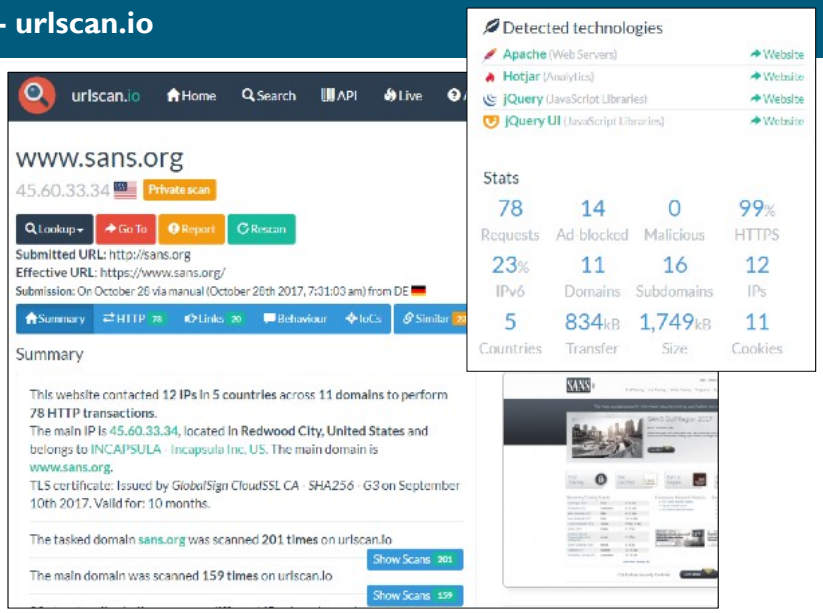
We can gather data from internet sites using many methods. While a good bit of the data moving to and from your browser is images and videos, much of the rest is text based. Some of the sites we visit may track us or host malware that can exploit our systems. To help us stay safe and access the information we need, we can use a variety of methods to retrieve web content.

- A proxy web application can retrieve the content from the remote system, analyze it, and then display the content. If the remote site has malware on it, your system doesn't get infected. Additionally, the web site does not know who asked the proxy site to retrieve the data, so your traffic is more anonymous.
- Command-line tools can be quick avenues to download content.
- Scripts can be written to retrieve and process remote data, making fast acquisition and analysis of large amounts of data possible.
- APIs (Application Programming Interfaces) allow for scripts (and people too) to retrieve data in structured formats (such as XML) instead of in HTML. This allows us to take the output and push it into an analysis engine.
- Cached content from search engines and other sites can help us retrieve data (especially older data) from sites without alerting the site we are acquiring the content.

Proxy Web Application - urlscan.io

Scans sites from the server and not your browser

Provides meaningful information about technologies used, links, subdomains, and network information



Proxy Web Application - urlscan.io

urlscan.io is a capable application that helps us understand what links, HTTPS certificates, images, and other technologies a web site has deployed by performing on-demand scans of the site. You provide a URL and, after analyzing the web site, urlscan.io displays a summary page and tabs across the top of the window that will highlight other data. Clicking through each can give an analyst an excellent picture of how that target site is implemented.

Each report is stored on urlscan.io's server and is retrievable via its unique URL. This allows you to run multiple scans of a web site over time and evaluate differences. As an example, https://sec487.info/9h was an evaluation of the sans.org web site run on October 28, 2017 and https://sec487.info/rh is another one performed September 4, 2019.

By asking urlscan.io to scan a target web site for you, you give away a little bit of your operational security because now a third-party web site knows that someone from your browser and your IP address is interested in a certain web site. We can minimize this risk by anonymizing our browser and using a VPN.

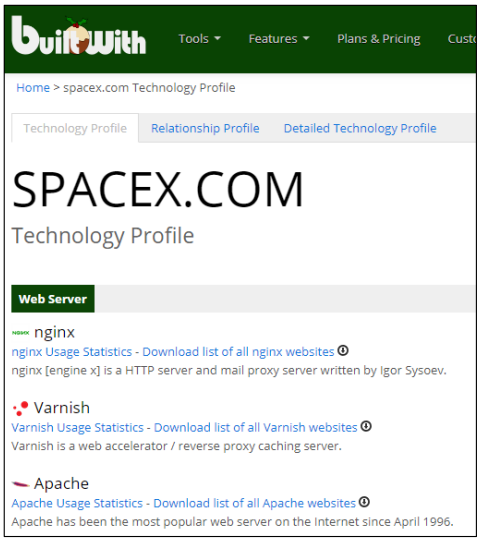
Images from https://sec487.info/9h, September 4, 2019.

Netcraft and BuiltWith

Two more sites that scan systems on the internet and examine the components they are built with:

- https://builtwith.com/spacex.com
- https://toolbar.netcraft.com/site_report?url=spacex.com

They examine response contents and provide you details



Netcraft and BuiltWith

Both BuiltWith and Netcraft web sites constantly scan the internet and extract data from the web sites they find. They show the pieces that make up the web page such as what technologies are used to make it, where the servers are, and data about SSL/TLS certificates.

For the Netcraft site, you will need to use the following format for queries: https://toolbar.netcraft.com/site_report?url=DOMAIN. So, if you wanted to research the SpaceX.com domain, the URL would be https://toolbar.netcraft.com/site_report?url=spacex.com (https://sec487.info/e6).

Image from https://builtwith.com/spacex.com, December 23, 2017.

Command-Line Tools

Pros of using command-line tools:

- Less risk of infection
- Recursive retrieval
- Fast
- Can use them remotely via SSH

Downside:

- Need to be configured to look more normal
- May be intimidating to use
- Format of data output sometimes suboptimal

Command-Line Tools

Our OSINT platform operating systems may contain tools that can help us retrieve data from web sites. While these applications allow for rapid interaction and content retrieval, they also can alert the destination web site that someone is doing something non-standard on the site. For instance, Wget (https://sec487.info/e7) and curl (https://sec487.info/l6) can make multiple requests to a web site and pull back resources (images, videos, text, etc.) to be stored locally on the analyst's computer. When performing these tasks, they make requests with user agent strings that give-away that they are not a normal web browser. Additionally, the speed at which they gather the remote content is much faster than a human using a web browser could. If the target web site has been configured to alert on these behaviors, the content you retrieve may not be accurate and your IP address may get flagged on the remote system.

Retrieving the Web Files Using Wget

- Installed by default on most macOS and Linux systems
- Retrieves files from web sites and saves them locally on your computer's file system

Switches of Interest

- `-e robots=off`
- `--no-check-certificate`
- `-A [extensions]`
- `-U "user agent"`
- `-r`
- `-l# (lowercase "L" and a number)`

Example: Retrieve only PDFs 3 levels deep from a site

```
$ wget -r -l3 -e robots=off --no-check-certificate -A .pdf
https://www.thirdpresbyterian.org/
```


Wget

This free Wget tool visits the web site you give it and retrieves all the linked and visible resources and stores them as local files on your computer. The files downloaded will be the content that you would normally see in your web browser and not the actual source code from the remote web site.

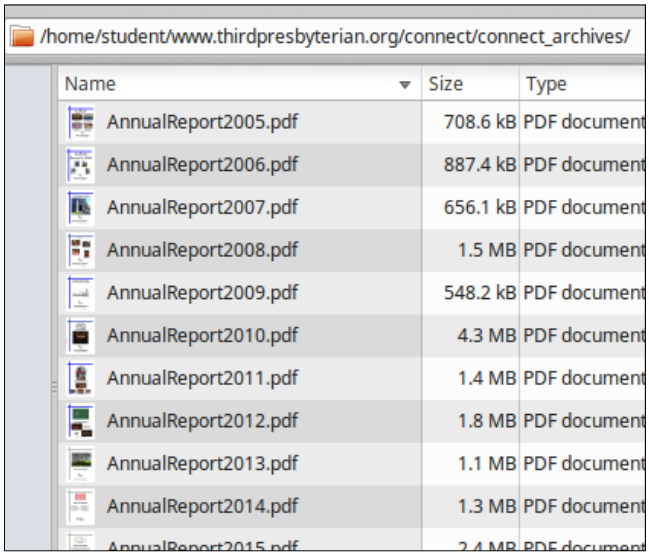
Wget has a variety of important flags and switches that we can pass to it when we execute it on the command line. While you can examine them for yourself by running "wget -h", we suggest that you consider using at least the following switches in your work.

- `-U` = Alter the user agent string for all of your requests. This is the web browser that Wget will tell the web server that you are using to make these requests. We suggest you visit a site such as https://sec487.info/9j to pick a user agent string that reflects a generic user of the site. Wget's default string is "Wget/[version]", which informs the remote web server that someone (you!) is trying to download all their data for some non-standard reason.
- `-r` = The "recursive" flag tells Wget to recursively gather data from the site and retrieve data from each directory and then its subdirectories too.
- `--no-check-certificate` = Some of the web sites you will want to recover data from will have HTTPS but invalid, expired, or self-signed certificates. This switch tells Wget to ignore those.
- `-e robots=off` = Wget will not spider a site if the robots.txt file prohibits it. This switch and value tell Wget to ignore robots.txt and grab the data anyway.
- -A .ext = This directs Wget to only download files with certain extensions. If you only want Microsoft Excel files, you could specify -A .XLS,.XLSX and wget will download those file type.

Wget will save the content it finds into the current directory that it is executed in unless you tell it to put it in another place. It will create directories and files that you can browse using the command line or a GUI tool.

Wget Results

- Files are saved in the directory structure from the web site
- Can use the "-nd" flag to put all files retrieved in a single directory, which may be helpful for later analysis



Wget Results

If we ran the command from the previous slide, we would see Wget retrieve many PDFs from the target web site and save them on your computer's local file system for further analysis. In this case, since we did not pass the "-nd" (no directories) option, the files are saved on our computer in a main directory named after the domain of the site. Then, each file downloaded is stored in the directory location where it would be found in the site. In the above slide, we see the "AnnualReport2008.pdf" file is stored in the "connect/connect_archives" directory path. These files would be found in the https://www.thirdpresbyterian.org/connect/connect_archives/ location on the remote web server.

curl

Unlike Wget, curl specializes in submitting and retrieving specific resources from sites instead of all of them

Use the "-A" switch to change the user agent string sent to the server

```
└─> $  curl -A "2018 Chevy Volt" https://www.sans.org > index.html
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 48176    0 48176    0     0  14619      0 --:--:--  0:00:03 --:--:-- 14620
```


curl

Since the curl tool is installed on most Unix, Linux, and macOS systems by default, it becomes a natural tool to use when trying to retrieve a few web elements. Need to pull an image, video, text, or script resource from a server? curl can assist, and it can also submit form data, log in to sites, and pass a wide variety of parameters and headers to the destination web site. Similar to Wget, curl retrieves the executed output of the web server content instead of gathering the web application source code itself.

curl sends the data it receives to the command line (standard out), so we may need to redirect it to save the content to a file using the ">" redirector. We demonstrate this in the above image by redirecting the output of the www.sans.org web page to an index.html file on our file system. Once we have this file on our system, we can launch a web browser and open the file.

There is an excellent, free online manual at https://sec487.info/k7, and the manual (man) pages are located at https://sec487.info/k6. Both will help you customize your curl usage.

Image from https://sec487.info/k8, October 28, 2017.

Finding Content via Built-in Tools

`grep` is free, built into macOS and Linux, and is a command-line/terminal tool for finding content

Can find strings or use regular expressions (regex) to find more complex patterns

Many switches that can be used to alter grep's behavior

On Windows computers grep can be installed:

- In the Cygwin application or the Linux Subsystem
- Via the GnuWin binaries

Alternatively, use "findstr" built-in command and alter your syntax or use PowerShell

Finding Content via Built-in Tools

We have a number of free tools that come with our modern-day operating systems that can help us find specific content within files we download. On macOS and Linux computers, the grep command-line tool is the most complete and useful application for our searching. We will explore how to use it to search through a file or a group of files to look for a certain string of characters or, using its regular expression capabilities, search for more complex patterns such as email addresses and phone numbers. grep has a large number of switches and flags that can be passed to it to change its searching or display behaviors. These options can be found on the manual (or "man") pages at https://sec487.info/l9.

Windows operating system users can choose to install grep using a variety of methods, such as through the Linux Subsystem, using the free Cygwin application (https://sec487.info/l7), or via the free GnuWin binaries (https://sec487.info/k4). Of course, Windows systems have their own native method of searching through files that analysts can use. From a command line, users can use the "findstr" (https://sec487.info/l8) application or leverage the Windows PowerShell scripting language.

pdftotext (Linux Command)

grep is terrific but cannot search well inside some of today's files without them being converted to a text file first

`pdftotext` converts PDFs to text files

To run on a directory of files, run:

```
for file in `ls *.pdf`; do
pdftotext $file; done
```

This bash loop lists all *.pdf files in a directory and runs them through the pdftotext application

When finished, the directory will have "pdf" and "txt" files with the same names

You can skip this step by using the `pdfgrep` tool

pdftotext (Linux Command)

While the grep tool is fabulous for finding strings within text files, it does not work well in some of the newer and more-common formats, such as PDFs and Microsoft Office documents. To use grep on non-text files, we may need to convert them into a format grep can use. The pdftotext tool (https://sec487.info/la) does just that: it converts PDF files into text so that grep and other tools can parse them.

The simplest form of the command is to type `pdftotext NAMEOFPDFFILE.pdf` in a terminal window, and the tool will extract the text content from the PDF and create a text file of a similar name to the original PDF. In the case above, the NAMEOFPDFFILE.txt file would be created.

To get pdftotext to run on a directory of PDF files, we must use a bash for loop to find each file with a ".pdf" file extension, convert it, and then continue on. That is what the bash loop in the slide above accomplishes.

The pdfgrep tool (https://sec487.info/le) searches within PDFs without the need to convert them.

Grep Results for "birth of" or "death of"

```
grep -A2 -B2 -i -E "(birth|death) of" *.txt
```

Using a regular expression (`-E`), we can quickly find our data

The `-A2 -B2` flags show us the 2 lines before and 2 lines after the match

```
messenger.txt-Lifting up Our Joys
messenger.txt-Birth
messenger.txt:For the birth of Easton Theodore
messenger.txt-Miller, to parents Lisa and Scott
messenger.txt-Miller, and grandparents Carol and
--
messenger.txt-
messenger.txt-Roderic and Marcia Frohman and
messenger.txt:family on the death of Rod's brother,
messenger.txt-Nathan, on November 20.
messenger.txt:Kathryn Thomas on the death of her
messenger.txt-sister, Beth Sinclair, on November 11.
messenger.txt-
```


Grep Results for "birth of" or "death of"

Once the files we have collected are in a format that grep can parse, we can start searching for content. Regular expressions, commonly called regexes, can be valuable timesavers when it comes to searching files for advanced strings, such as credit card numbers, cryptocurrency wallet IDs, and unique strings of characters. There are many online tools (https://sec487.info/lb) for learning regular expressions. We will dive into a few of them in the coming section and how the grep tool can help us find that content.

In the slide above, we search through all the files with a ".txt" file extension and use a regex (`-E` switch) that will look for the word "birth" or the word "death" in combination with the string " of". The "`-i`" flag tells grep to match the strings regardless of the upper- or lowercase target content. The "`-A2 -B2`" values ask grep to retrieve the two lines after and the two lines before the matched string so that we can get more context of where the string was used in the document.

Grep Results for Email Addresses

```
grep -o -h -E "\b[a-zA-Z0-9.-]+@[a-zA-Z0-9.-]+\.[a-zA-Z0-9.-]+\b" *.txt | sort -i -u
```

Here we use a regex for email addresses and ask grep to only (`-o`) show the matches and not the filenames (`-h`)

We then use the sort command to remove duplicates

Always examine the data for errors

```
sally6913@aol.com
sandygia@aol.com
saraheboyce@hotmail.com
sarah@wisbeydesign.com
scarter4@earthlink.net
sdaniel5@rochester.rr.com
selliott@rotoliteelliott.com
Selliott@rotoliteelliott.com
shellman@rochester.rr.com
Shellman@rochester.rr.com
silvia.widish@gswny.org
sjdanielson@aol.com
skieren4@yahoo.com
slocke43@rochester.rr.com
slocke43@rochester.rrocom
slocke43@rohester.rr.com
```


Grep Results for Email Addresses

Switching to look for email addresses in text files, we can use the grep command and a regex similar to the one above (from the https://sec487.info/k5 site). As can be seen in the results image on the right of the slide, this is not a perfect method of extracting email addresses from documents. There are duplicates due to case (upper and lower) and errors from the pdftotext conversion. However, this technique is extremely fast and parsed 731 text files in 0.038 seconds.

Grep Results for US Phone Numbers

```
grep -o -h -E "\(?[0-9]{3}\)? ?[0-9]{3}[-\.]?[0-9]{4}" *.txt | sort -u
```

We can create regexes for US phone numbers as well

Keep in mind that phone numbers can be found in a variety of formats, as shown in the output of the command on the right

Always examine the output for errors

```
(202) 224-3121
2507786325
(347)668-6181
388271-6513
(585) 2542697
(585) 260-4379
(585)2716513
(585) 2716513
(585) 271-6513
(585) 271.6513
(585)271-6513
585271-6513
(585) 271-6537
585271-6537
```


Grep Results for US Phone Numbers

We can search for whatever content we wish in files. Above, we chose to match United States phone numbers using a regex. When creating the regexes, keep in mind that the content you wish to match may be in a variety of formats. As shown in the output, we have phone numbers separated using "-" and using ".". Area and country codes may also be in a variety of formats. Make sure you examine the target documents prior to creating the regex to ensure you do not make a regex too narrow in scope.

Grep Results for International Phone Numbers

International phone numbers are complicated as they can be grouped differently, have country codes, and have different delimiters

We can check regexes against sample data on websites like regextester.com

Grep Results for International Phone Numbers

Just as we can search for phone numbers from the United States, we can do the same for any string of characters, including international phone numbers. These numbers usually have country codes (for example, +31) and can have a variety of different groupings of the numbers. To take all of this into account, our regular expression gets extremely complicated. There is one that appears to work well that we found on the regextester.com website (https://sec487.info/q8). The regex looks like what is shown below and, in the slide above, is found at arrow 1. The site also allows users to submit sample data to see if the regex will match. We see above (arrow 2) that some of the strings matched (they are highlighted blue) and a few at the bottom did not.

One regex, from the above site, for international phone numbers is:

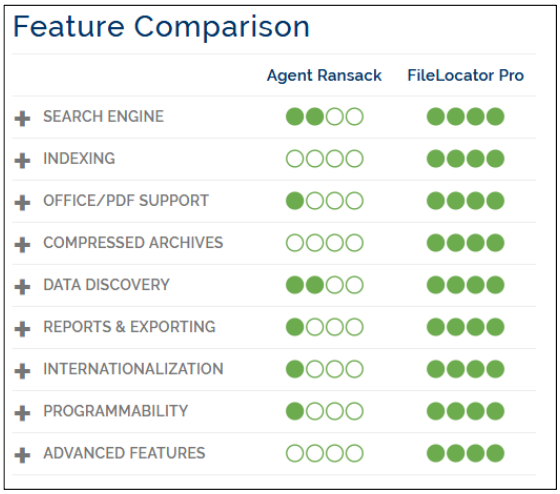
```
((?:\+|00)[17](?: |\-)?|(?:\+|00)[1-9]\d{0,2}(?: |\-)?|(?:\+|00)1\-\d{3}(?:
|\-)?)?(0\d|\([0-9]{3}\)|[1-9]{0,3})(?:((?: |\-)[0-9]{2}){4}|((?:[0-
9]{2}){4})|((?: |\-)[0-9]{3}(?: |\-)[0-9]{4})|([0-9]{7}))
```

Mythicsoft Tools

Agent Ransack (Free) and FileLocatorPro (~$60) are excellent graphical tools that work similarly to grep but on Windows computers

Simplifies searching of files

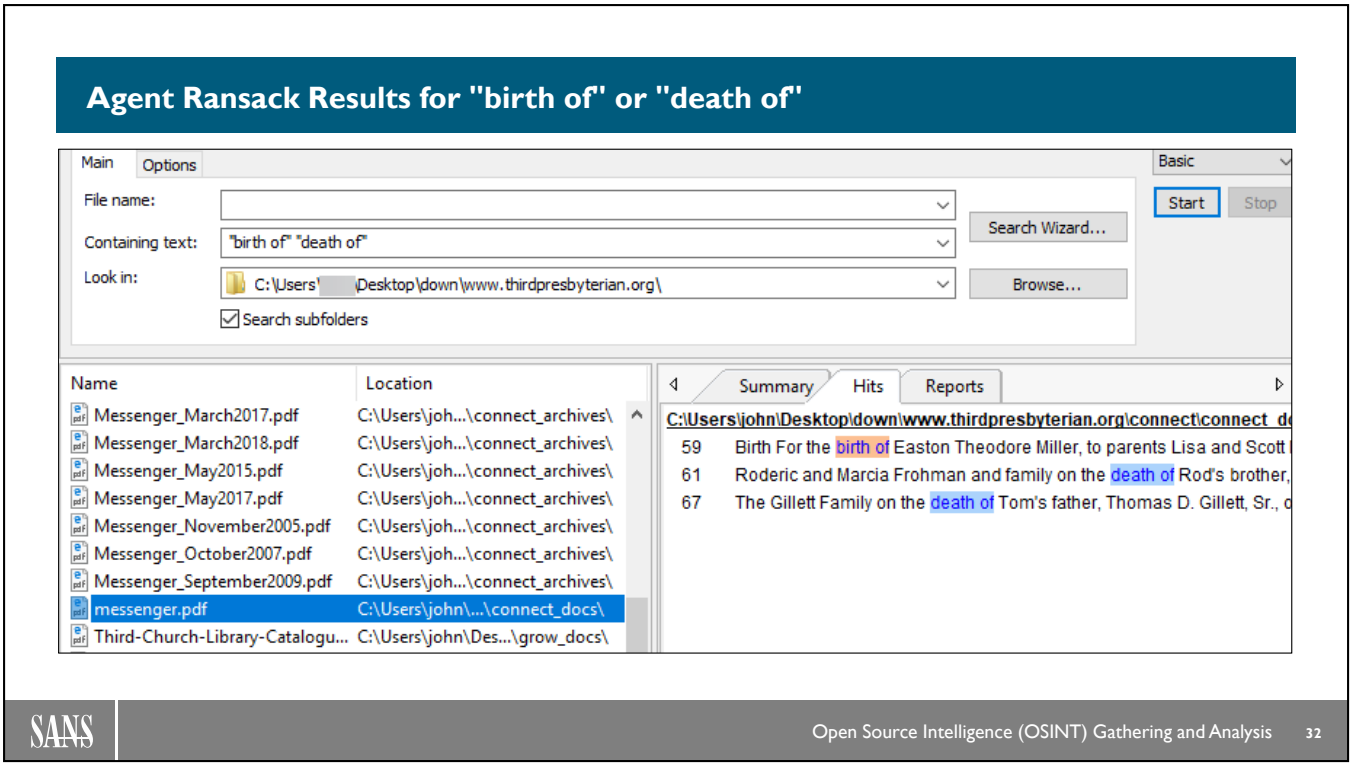
Can use regexes!



Mythicsoft Tools

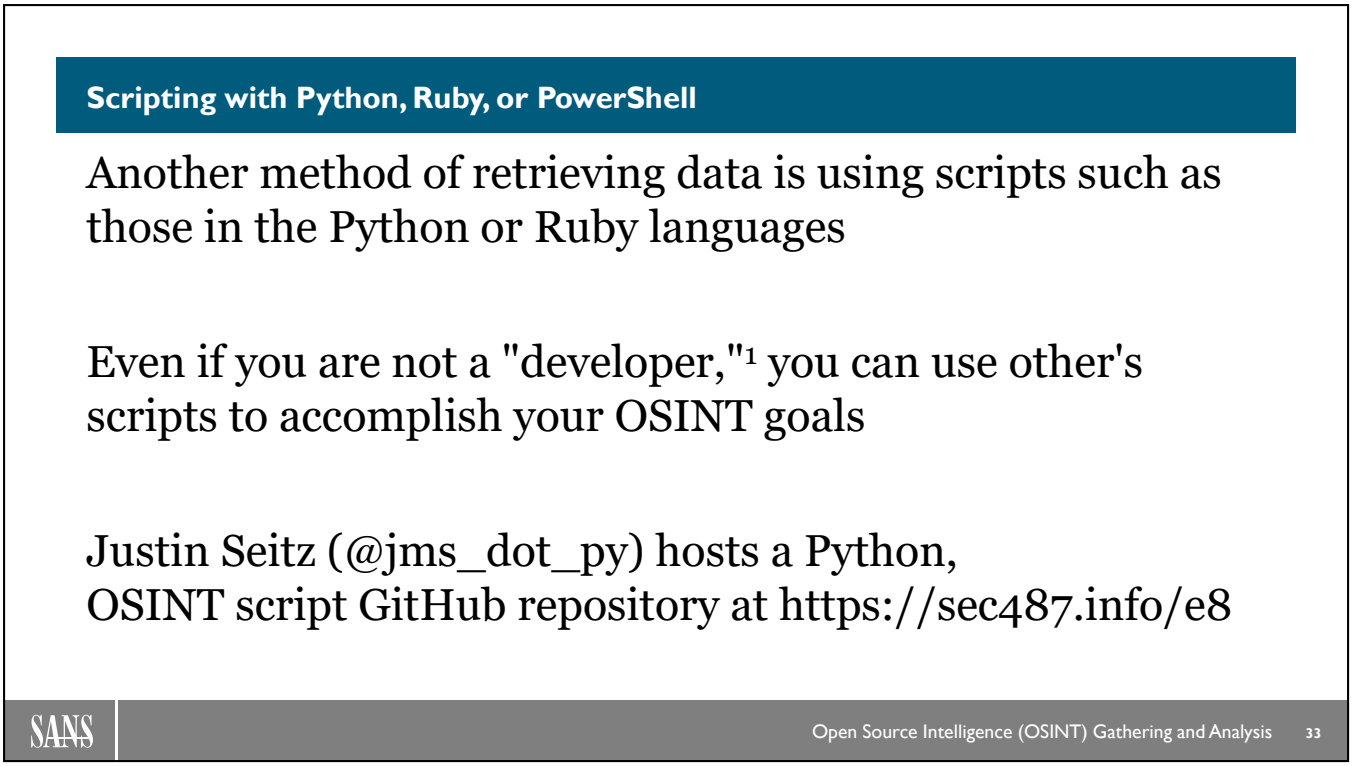
There are two excellent file content searching tools that are like grep but much easier to use. They do not retrieve files (you will need to use Wget or another tool for that) but, instead, allow you to quickly search the content of a wide range of file types for specific strings and regexes. There are two versions of the Mythicsoft tools: the free Agent Ransack and FileLocatorPro (which costs around $60).

Image from https://sec487.info/lc, December 31, 2018.



Agent Ransack Results for "birth of" or "death of"

Using the Agent Ransack application to run a similar query to the grep query we ran earlier, we see the graphical interface and the same results we found before. This time, however, we did not have to convert the files into a text format to extract data, and the tool has an easy-to-use graphical interface.



Scripting with Python, Ruby, or PowerShell
In recent years, scripting languages like Python (https://www.python.org/) and Ruby (https://www.ruby-lang.org/) have become easy-to-use tools for gathering and manipulating our assessment data. These scripting languages allow us to retrieve data from web sites and analyze that and other data in efficient manners. If you already know one language, use it. You don't have to learn Python or Ruby. Python is a great starter language if you do not already have scripting skills, as there are excellent online tutorials and eBooks, and the community has published hints, tips, and tricks to the internet. Many of our OSINT scripts and tools are written in Python too!

Python and Ruby are free, interpreted languages that are either installed by default on many operating systems or can simply be downloaded and installed.

To give you an idea of how simple it is to use Python to retrieve a web page, the following four lines of code, typed in a terminal window, will retrieve the https://www.sans.org web page and display it on the screen (just like curl did!):

```
$ python3
>>> import requests
>>> r = requests.get('https://www.sans.org', verify=False)
>>> print(r.text)
```

Justin Seitz (@jms_dot_py) has an amazing set of free python tools in his public GitHub repository at https://sec487.info/e8

Reference:

[1] We hear people say "I'm not a developer. I'm not a coder. I cannot make a script." all the time, and it simply is not true. Scripting is about writing or altering existing code to meet your needs. Search the internet for the terms "python" and "osint" and you will see many responses. Consider using a Python tutorial to gain a few scripting skills and then try to alter some of these existing scripts to achieve your goals.

EyeWitness Python Tool

Free tool from Chris Truncer

Give it a list of hosts and protocol (web, RDP, VNC) and it visits them

Records screenshots and response header data

```
student@sec487 (11:58:54) :/opt/tools/EyeWitness$ ./EyeWitness.py -f ~/urls --web

################################################################################
#                                  EyeWitness                                  #
################################################################################
#           FortyNorth Security - https://www.fortynorthsecurity.com           #
################################################################################

Starting Web Requests (5 Hosts)
Attempting to screenshot http://osintcurio.us
Attempting to screenshot http://sans.org/sec487
Attempting to screenshot http://webbreacher.com
Attempting to screenshot http://osintframework.com
[*] Hit timeout limit when connecting to http://sans.org/sec487, r    g
Attempting to screenshot http://osintframework.de
Finished in 30.7278749943 seconds

[*] Done! Report written in the /opt/tools/EyeWitness/10052019_115901 folder!
Would you like to open the report now? [Y/n] y
```


EyeWitness Python Tool

Chris Truncer's company Forty North Security has a GitHub repository with the EyeWitness web site tool (https://sec487.info/yf). This free Python command-line tool accepts lists of hosts for it to visit. While we mainly have used it to visit web sites, and thus pass it lists of web sites, the tool can also visit systems with Remote Desktop Protocol (RDP) and Virtual Network Computing (VNC) ports that are open.

What EyeWitness does—and why we might add it to our list of OSINT tools—is it takes your list of systems and visits each one. When it interacts with the computer, it takes a screenshot of whatever is displayed and then it captures other response data. For instance, if it visited an RDP system that had a Windows login page, it would capture the Windows login profile icons and the user names and domains that were displayed. Then it compiles all of these responses into an HTML report for you to browse. It vastly speeds up the process of investigating multiple systems.

In the example above, we ran the tool (1) and fed it a list of domains (2), telling the tool that these domains were web sites (1). As shown, EyeWitness attempted to visit each web site (2) and take screenshots of what it "saw." At the end of its run (3), it saved an HTML report to the file system and asked if we wanted to view the report. When we pressed "y" and hit Enter, it showed us what is on the following slide.

Chris Truncer hosts a blog post about the EyeWitness tool at https://sec487.info/yg.

EyeWitness Report

This is one section of the HTML report we are viewing in a web browser

Details on the left from the HTTP response and screenshot on the right

EyeWitness Report

The tool saves its output in HTML files locally on your system. The files contain one or more pages of HTML tables. In our example, we asked EyeWitness to visit several web sites, so on the left in the report (1) we have the HTTP response headers. On the right (2) is a screenshot of what EyeWitness was presented from the web server. EyeWitness is not great at talking screenshots of JavaScript-heavy web sites that are complex. The simpler the web site, the better and more complete the image taken by this tool.

Above, we see HTML tags like the page title (3) and can view what cookies (4) would have been set in our web browser had we visited the site. The HTTP responses may also reveal security protections the sites employ (like the HSTS code at 5) and other header data.

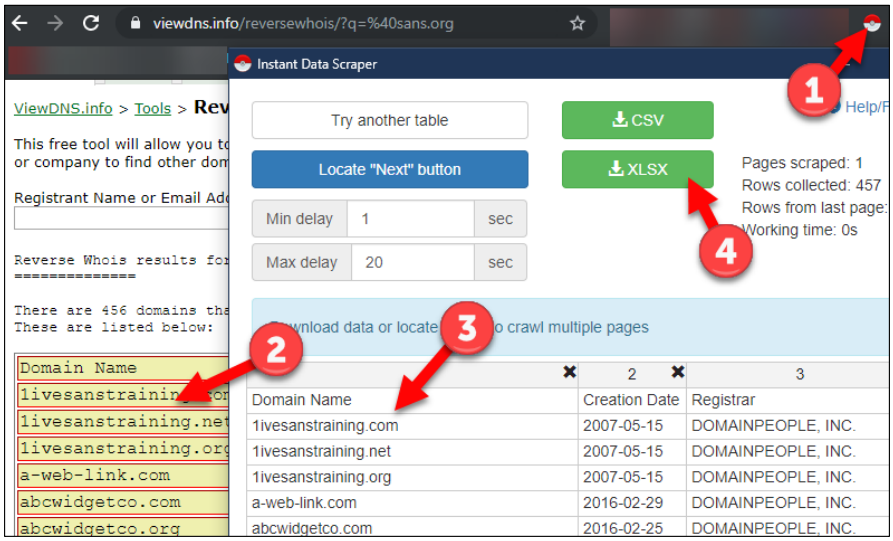
Browser Extension That Extracts Data

Instant Data Scraper
Offered by: webrobots.io

Have browser data in a table format?

Try the Instant Data Scraper extension for Google Chrome

Exports to CSV and XLSX



Browser Extension That Extracts Data

Many times, we will be in an OSINT investigation and, on screen, we will see data in a table format. We then try to capture that content for later use by selecting it and pasting into a document or by using the "Save As…" browser feature to download the whole page. Then we have to sort through the HTML code and extract the content we want. All of this is time consuming.

The Instant Data Scraper (https://sec487.info/r7) extension simplifies this process. Add it to your Google Chrome browser and then, when you browse to a page that has data in a table format you wish to capture, click on the extension icon ("1" in the image above). The extension will analyze the page and try to determine what it can extract. Then it will show what it thinks you want extracted (2). After that, in a new window, it shows what a CSV or XLSX file would look like with that content (3), and presents you with the buttons to perform the export (4).

Images from https://sec487.info/r7 and https://sec487.info/r8, September 4, 2019.

APIs

Application Programming Interfaces (APIs) are methods of retrieving data in computer-friendly formats

Web site content sent to our web browsers is made for human presentation with images and scripts to "pretty it up" for us

APIs allow scripts to get web site content in formats that they can easily consume

Common API output is in XML or JSON formats

Sometimes APIs allow us access to data not retrievable via a web page[2]

APIs

When we browse web pages, our browsers retrieve JavaScript, images, HTML, videos, and content that render, execute, and display in our browsers. Browsers assemble the pieces into the correct places and execute functions that make the web site content easier for humans to interact with. We can use scripts and automated tools to pull similar data from web sites using application programming interfaces[1] (APIs) but in XML (Extensible Markup Language) and JSON (JavaScript Object Notation) formats, without images and fancy JavaScript. XML and JSON are formats that organize data for scripts to process.

Some APIs require users to register with the web site and obtain an API key (just a unique string of letters and numbers that identifies the user) that must be presented to the site before it will send data. Other APIs are free to use without keys.

APIs are useful to us when retrieving large amounts of data from web sites. Sites returning a single result to our searches are simple to process for our OSINT needs. When hundreds or thousands of records are returned, we will need a less-manual method of parsing and analyzing that content. APIs help us by formatting the data responses into structured sections that other applications and scripts can easily parse.

Additionally, as shown in the OSINT Curious Project's 10-Minute Tip video "Using APIs to Reveal Data,"[2] by requesting data from APIs and in other formats than what is rendered in the web browser, we can many times get access to data that we would not have seen.

Reference:

[1] https://sec487.info/9l

[2] https://sec487.info/q7

API Output Example: api.opencorporates.com

Human-Readable

AXE UNILEVER

Company Number	531289130
Incorporation Date	15 April 2011 (over 8 years ago)
Company Type	Société à responsabilité limitée (sans autre indication)
Jurisdiction	France
Registered Address	55 RUE D'ANJOU PARIS 8 75008 PARIS FRANCE
Industry Codes	56.10C: Restauration de type rapide (Nomenclature d'activités française (2008)) 56.10: Restaurants and mobile food service activities (European Community NACE Rev 2) 5610: Restaurants and mobile food service activities (UN ISIC Rev 4)

API JSON Output

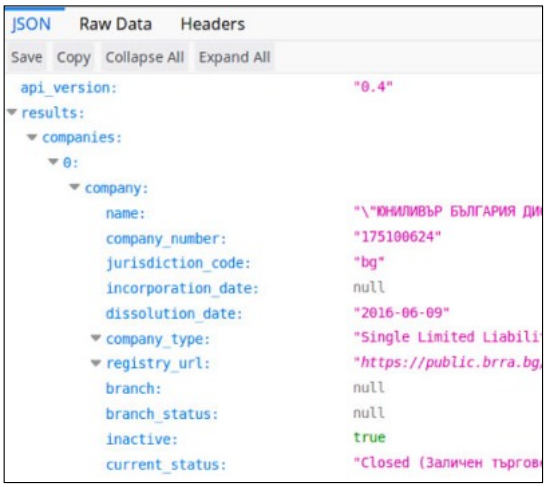
API Output

APIs are usually found at distinct hosts (such as http://api.example.com) or paths in the application (such as http://example.com/api/). A great OSINT example of an API is found at https://sec487.info/9o, shown in the right image above, versus the human-readable site on the left (https://sec487.info/9m).

Looking at the details of the AXE UNILEVER company on the left shows us some details that are also apparent in the image on the right. The "AXE UNILEVER" company name and the incorporation date appear in both images but are formatted differently. But in the right image, we can see that the dates and times when this record was created, updated, and retrieved are all displayed in the JSON from the API but not in the other page. We will see this occur repeatedly throughout the course when we examine API and JSON content. We simply get more information than what we could get only by viewing the web page in the browser.

Default View of JSON in Web Browsers

Firefox



Chrome

Default View of JSON in Web Browsers

Most modern web browsers can display JSON. Some, however, display it better for the human eye to read. Viewing the JSON at https://sec487.info/9o in both Mozilla Firefox (on the left above) and Google Chrome (right), we can see that Firefox, by default, renders this JSON in human-readable or "prettified" format (sometimes also referred to as "beautified"). Chrome displays the raw JSON content. Both show the exact same content just in different formats.

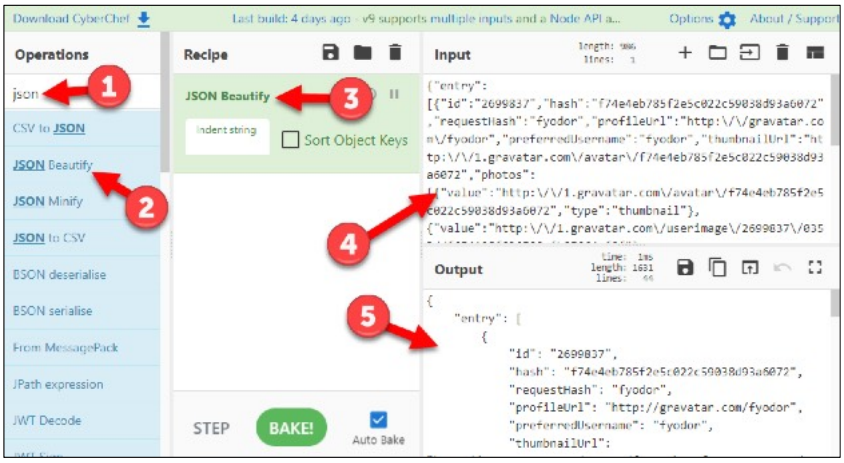
To make the JSON easier to view within Google Chrome, you can install Chrome Extensions that will beautify JSON. One such extension is the JSON Formatter (https://sec487.info/r0). Once that is installed in Chrome, the format of the JSON shifts to the much more readable format shown below.

CyberChef: The Master Formatter and Extractor

CyberChef, by GCHQ, is a free data converter, extractor, hasher, and more!

Download it or use online

At right, we show how to beautify JSON



CyberChef: The Master Formatter and Extractor

The Government Communications Headquarters (GCHQ) created and released a free, amazing tools that will serve most of our data transformation and extraction needs. CyberChef (https://sec487.info/r2) is both an online and offline tool. This is an amazing site that is invaluable to OSINT investigators. One reason why you might use it is to transform JSON content into a human-readable format, as shown in the slide above. Following the numbers in the diagram we search for the term json (1), select "JSON Beautify" (2), and drag it into the "Recipe" pane (3). Next, we paste the JSON we want to parse in the "Input" pane (4) and read the output (5).

CyberChef does many other things that we might find useful, including:

- Extracting email and/or IP addresses from text
- Base64 encoding and decoding
- Creating hashes (MD5, SHA, etc.)
- Converting UNIX or Epoch time to other formats
- Extracting EXIF metadata from files

This tool also allows us to create recipes where we use multiple operations to achieve our goals. So if you have a string that is URL encoded and then Base64 encoded, you can create a recipe to decode each of those in a single step. Common recipes can be saved on your system for later use.

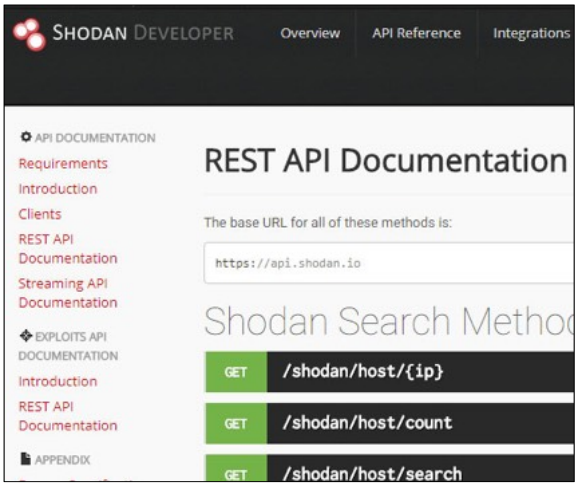
Lastly, if you need to use CyberChef but don't want to use it on the web site, you can download it and move it off the internet. For additional details and usage examples, visit The OSINT Curious Project's 10-MinuteTip at https://sec487.info/r3.

Reference and image from https://sec487.info/r1, September 3, 2019.

API Documentation

To use some APIs, you need to understand what commands to send them

Most sites will have documentation with examples of what commands should be used and how



API Documentation

To interact with a web site's API, we need to understand what commands it accepts, in what format, and with what authentication. All of these can be found in the API documentation usually found on the site's web servers. Sometimes it is at a different URL/domain from the main web site, as the example above from https://sec487.info/e9 illustrates.

With a little bit of reading, we can sometimes use the APIs to gather large amounts of data fast. APIs are not limited to just retrieving data, though. Depending upon the site, you may be able to tweet as the currently logged-in user, post a photo, delete a calendar entry, or follow a friend.

Our scripts and applications use these API commands to interact with the API and achieve our OSINT goals.

Image from https://sec487.info/e9, October 28, 2017.

Public API Collection

Huge list of free, public APIs available in a GitHub repository

Many require authentication and API keys

Many categories of content, from geocoding to sports & fitness

Social

API	Description	Auth	HTTPS
Buffer	Access to pending and sent updates in Buffer	`OAuth`	Yes
Cisco Spark	Team Collaboration Software	`OAuth`	Yes
Discord	Make bots for Discord, integrate Discord onto an external platform	`OAuth`	Yes
Disqus	Communicate with Disqus data	`OAuth`	Yes
Facebook	Facebook Login, Share on FB, Social Plugins, Analytics and more	`OAuth`	Yes
Foursquare	Interact with Foursquare users and places (geolocation-based checkins, photos, tips, events, etc)	`OAuth`	Yes

Public API Collection

To discover other APIs that may be of use in your OSINT work, visit the Public APIs project on GitHub (https://sec487.info/lf). They have documented and categorized a large number of public APIs that you may useful. The site provides hyperlinks to the web applications with the APIs and documentation and lists some basic information about each API.

Cached Content

Some web applications scan the internet and preserve copies of what they find

They allow us to examine sites "offline" and go back in time to see what data a site showed months or years before

Fully cached content sites include:

- Archive.is
- Archive.org

Search engines cache some content

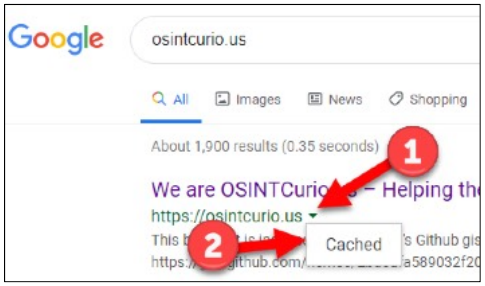
Cached Content

Most web spiders for search engines scan the internet, find web pages, collect data, and allow people to search for that data. Many of those engines also save the data found for specific sites and allow users to retrieve cached copies of that site. Some other web spiders make copies of what sites looked like when they were scanned. Their purpose is preservation of the content and look of the site.

The Archive.org and Archive.is web sites cache a full copy of a web site, including scripts and images. Using web applications such as these allows us to view site contents in a stealthy manner and also go back in time to see what the site contents looked like.

Search Engine Cached Content

Most search engines allow users to retrieve cached copies of what they collected from a web site



Search Engine Cached Content

Search engines like Yandex, Bing, and Google scan the internet and index web site content. They store vast amounts of HTML content that they allow users to search through to find meaningful results. Using features on their web pages, users can select to view those cached pages that the search engines have retrieved instead of the users visiting the live web site of their target.

In the above images, if a user clicks the triangles next to the results domains (arrow 1 in both images), they can then select to see the cached copy of the website (arrow 2 in both images).

Images https://sec487.info/qo and https://sec487.info/qp, September 3, 2019.

Cached Content OPSEC

OPSEC WARNING - Search engines (Google, Yandex, etc.) pull some resources from the live web site

To prevent Google from pulling live resources, add **&strip=1** to end of URL

https://webcache.googleusercontent.com/search?q=cache:q0ser3Y903YJ:https://www.sans.org/course/open-source-intelligence-gathering+&cd=1&hl=en&ct=clnk&gl=us**&strip=1**

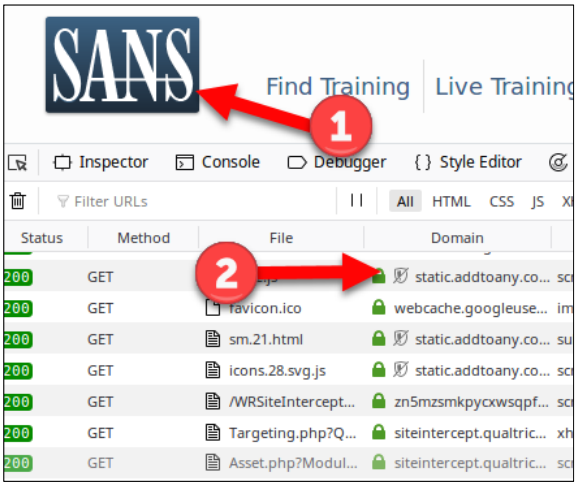
Cached Content OPSEC

When you use search engines to try to view a web site's content, it often will make your web browser reach out to the live web site and pull data such as images, JavaScript, and other web page parts. If your goal is to not let the site know that you are researching it, add "&strip=1" to the end of the Google cache URL, and Google will not load content from the original site.

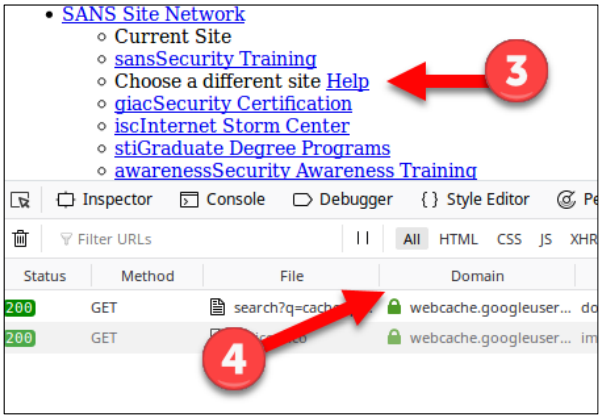
The example above pulls content from Google's cache and does not load remote content from the sans.org domain. A short URL to that link is https://sec487.info/iu.

Google Cache Results

Without "&strip=1"



With "&strip=1"



Google Cache Results

In the above images, we can see that when we did not use the "&strip=1" parameter and value in the Google URL, our browser retrieved images (arrow 1) from web sites outside of Google's cache (arrow 2). In the image on the right, we see that, by using the "&strip=1" value, the content retrieved is text-only (arrow 3) and limited to what Google had cached (arrow 4).

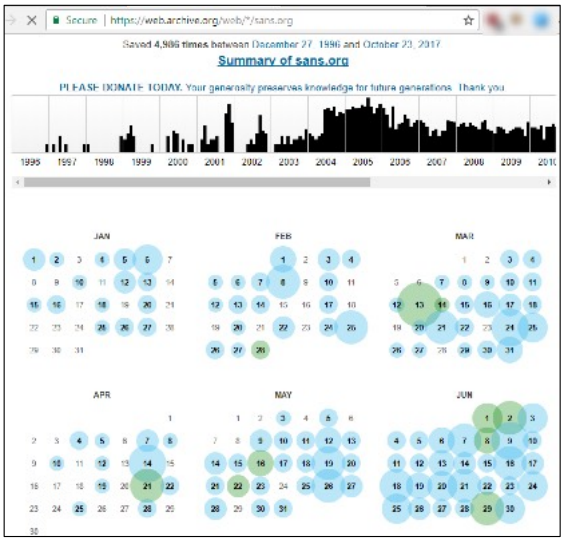
For more information about this, watch The OSINT Curious Project's 10-Minute Tip at https://sec487.info/qg.

Archive.org

Non-profit founded in 1996 that archives web site data, music, software, and more!

Maintains copies of cached content across years

Can use tools to grab all previous versions (see notes)



Archive.org

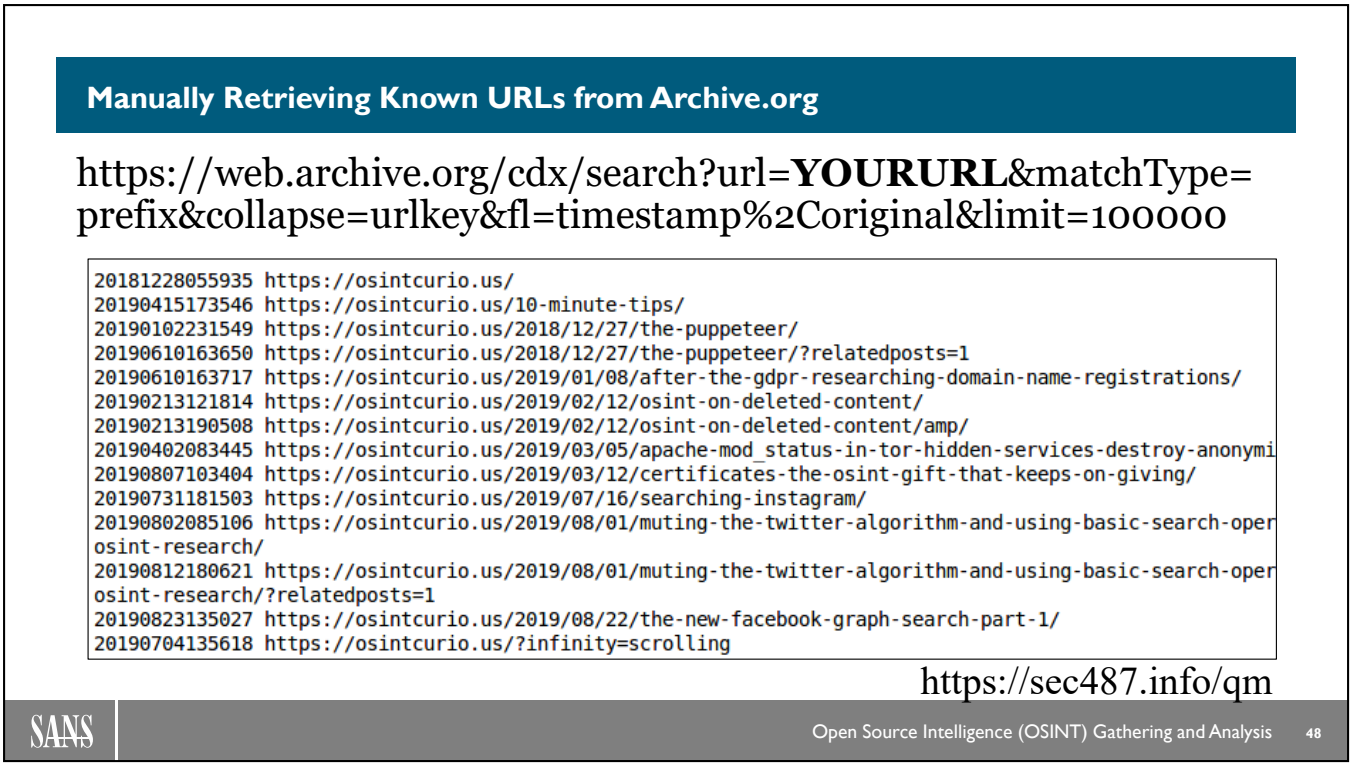
The Archive.org site states its purpose well:

"The Internet Archive is a 501(c)(3) non-profit library. Founded in 1996, our mission is to provide Universal Access to All Knowledge. We collect published works and make them available in digital formats. We are building a public library that can serve anyone in the world with access to the Internet.

“We began in 1996 by archiving the Internet itself, a medium that was just beginning to grow in use. Like newspapers, the content published on the web was ephemeral - but unlike newspapers, no one was saving it. Today we have 20+ years of web history accessible through the Wayback Machine." (https://sec487.info/k9).

In short, the site creates dated backups of web sites and makes that content accessible to us all. Since this data is 100% cached content, the target web site you are researching is not alerted that you are viewing its content. The Waybackpack (https://sec487.info/9k) contains code to pull down all the known archive.org archives for a given domain.

Image from https://sec487.info/9g, October 27, 2017.



Manually Retrieving Known URLs from Archive.org

Tools like the "waybackMachine" from https://sec487.info/qn accept a domain and then query the archive.org web site to find any URLs that have been indexed from that site. We can extract the URL that it uses and modify it a little bit to make a request to archive.org for the timestamp (when a web page was indexed) and the URL that was archived.

This is what we did in the above slide. We replaced the text "YOURURL" with "osintcurio.us" and submitted the URL in our web browser. The results have 2 columns: timestamp and URL archived. If any of those URLs looked interesting to us, we could visit archive.org and retrieve them.

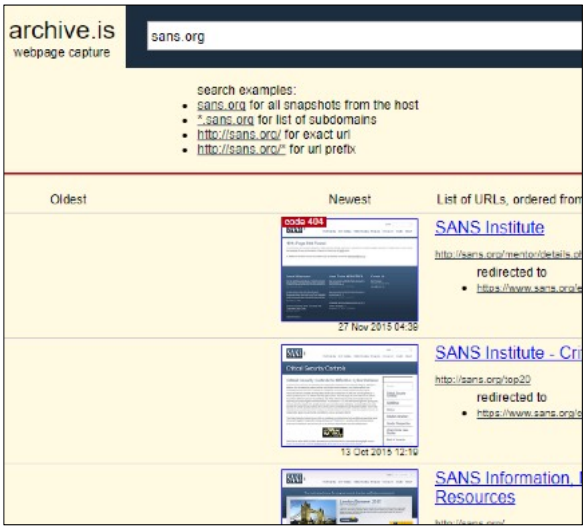
Image from https://sec487.info/qm, August 23, 2019.

Archive.is

Privately funded site that archives web sites (text, images, JavaScript)

Maintains copies of cached content across years

Fewer backups than archive.org



Archive.is

Another caching site is https://archive.is/, which is a privately funded site that makes full backups of web sites, their text, images, and JavaScript. It maintains these archives across years and allows users to retrieve those older copies of sites.

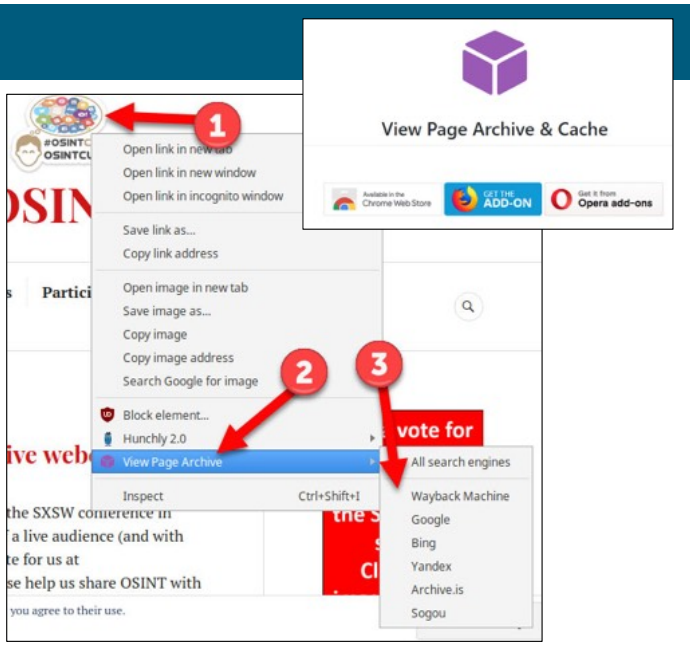
Reference and image from https://sec487.info/ea, October 27, 2017.

Browser Extensions for Archives

Armin Sebastian has web archive search extensions/ addons for several browsers that make searching multiple archive sites easier

Add them to your browser(s)

Right-click on image (1) and choose the image search engines to use (2 and 3)



Browser Extensions for Archives

As noted above, Armin Sebastian released several browser addons/extensions for Chrome, Firefox, and Opera that allow you to quickly search for an image in multiple image search engines. Just right-click on an image and choose where you would like to search. The extension will open a tab or tabs in your browser for each site you chose to search.

Reference and image from https://sec487.info/qk, August 23, 2019.

Google's Mobile-Friendly Test (1)

Another way we can retrieve a web page might also bypass **authentication and access controls**

We ask **Google's Mobile-Friendly Test** site to "check" a site[2]

It **retrieves the HTML, using Google IP addresses**

Sometimes sites let Google index places that normal users cannot

We ask Google to:

1. Check the site
2. Retrieve the HTML data
3. We save the HTML
4. Then we view the HTML code and can **"see" the web page**

Google's Mobile-Friendly Test (1)

Google has a web page dedicated to checking whether web sites are mobile-friendly (i.e., do they look good when viewed on your mobile device). The Express Tricks blog[1] posted a technique to get Google to pull content for you! Google's mobile-friendly test site (https://sec487.info/ww) works as follows:

1. You submit the URL you wish to examine to Google's web site.
2. Google, from its servers and IP addresses, visits the site and saves the HTML content.
3. Google then provides that HTML content to your browser.
4. You save the HTML into an HTML file or view it in an online resource.

There are benefits to using this approach in that this works without authentication. Also, Google visits the site from its servers and IP space, and sometimes web sites like Google visiting their site and will allow it deeper into the site than a normal user could go without authenticating.

This technique does not work on all sites. If the web page you want to retrieve requires authentication or has a restrictive robots.txt file (preventing Google from accessing content), this technique will fail.

References:

[1] https://sec487.info/wx

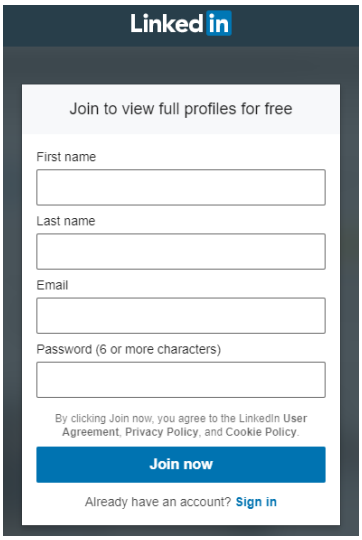
[2] https://sec487.info/ww

Google's Mobile-Friendly Test (2)

1. Find a URL to retrieve content from (LinkedIn.com requires authentication for normal users to access content)

`https://www.linkedin.com/in/natashasrulowitz`

Visit directly and get the "authwall" authentication page



Google's Mobile-Friendly Test (2)

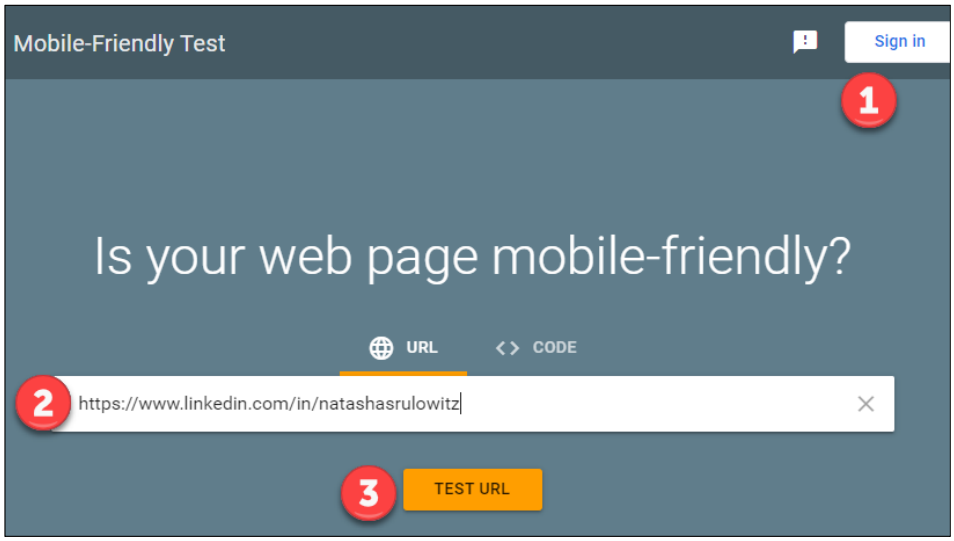
First thing we need to do is to find a URL that we would like to test this technique with. LinkedIn.com offers a great target, as this site requires us to authenticate as a valid LinkedIn user before we can access its data. We will use the random user profile at https://www.linkedin.com/in/natashasrulowitz for this test.

Visiting that URL directly redirected our browser to the LinkedIn authentication page and did not show us Natasha's profile.

Image from https://sec487.info/wy, October 3, 2019.

Google's Mobile-Friendly Test (3)

2. Visit Google's mobile-friendly page and paste in the URL (2)
3. Click TEST URL and solve the CAPTCHA



Google's Mobile-Friendly Test (3)

Next we head over to Google's Mobile-Friendly test site and paste in the target URL into the search field (2). Notice we are not authenticated to Google (1), nor are we logged into LinkedIn.com.

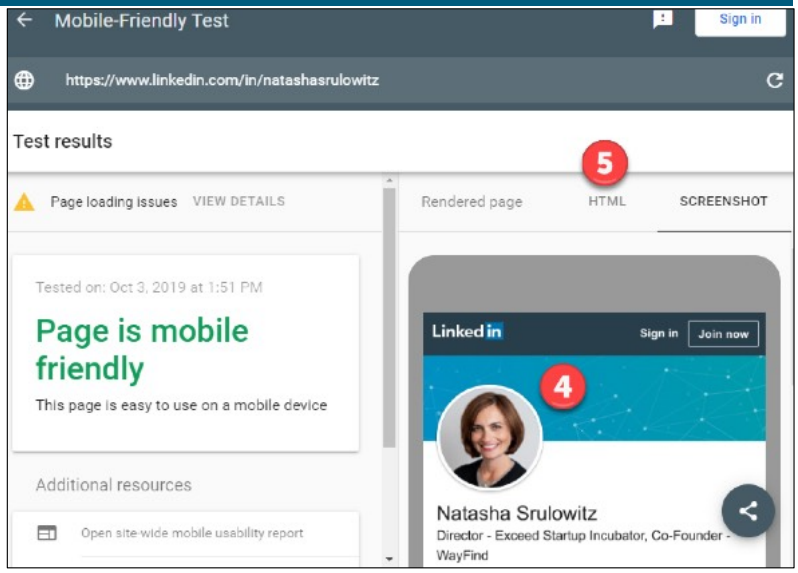
Clicking the TEST URL button (3) will most likely launch a CAPTCHA for you to solve (not shown above).

Image from https://sec487.info/ww, October 3, 2019.

Google's Mobile-Friendly Test (4)

4. Google retrieves the HTML and images but only shows some of it

5. Select the HTML tab



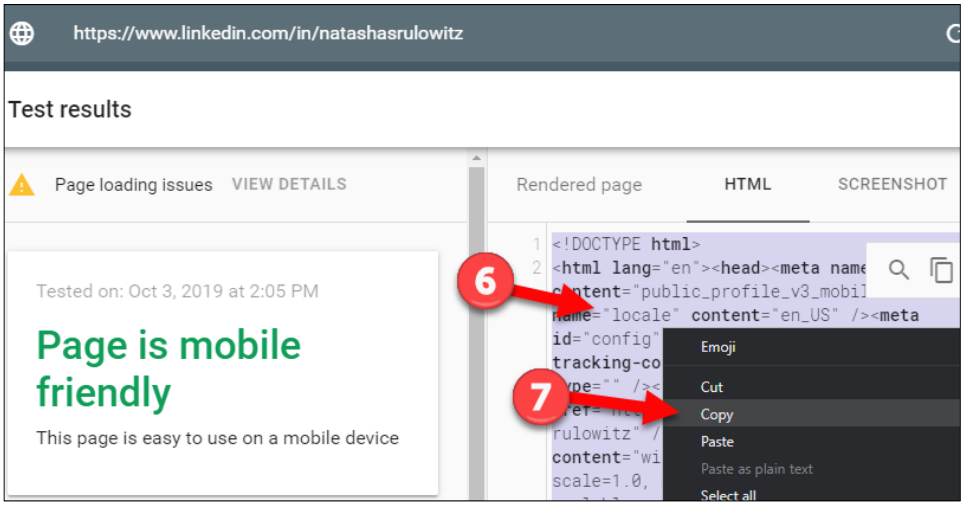
Google's Mobile-Friendly Test (4)

Google retrieves the content we asked for and shows it (4) on a mobile device mock-up (after all, we asked it to do a mobile-friendly test, didn't we?). We want to see all the content it retrieved from the target site. So we click the HTML tab (5), which reveals the HTML Google pulled back.

Image from https://sec487.info/wz, October 3, 2019.

Google's Mobile-Friendly Test (5)

6. Select all the HTML
7. Copy it to your computer's clipboard



Google's Mobile-Friendly Test (5)

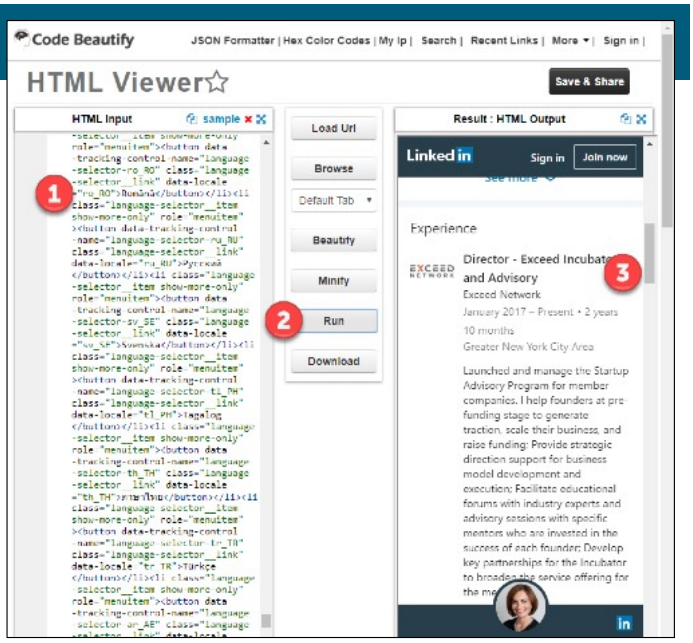
Select all the text inside the frame under the HTML tab (6) and copy it (7) to the computer clipboard/memory.

Image from https://sec487.info/wz, October 3, 2019.

Google's Mobile-Friendly Test (6)

For the last part, you can save that content into a local HTML file and then open it in a web browser

Or use a site like https://codebeautify.org/htmlviewer/ to paste (1), run (2), and view (3) it in a web page (on the right)



Google's Mobile-Friendly Test (6)

The final step in this process is to put that HTML content in your computer's memory somewhere.

The most secure and our preferred method is to open a text editor on your computer, paste the HTML contents in, and then save it as an HTM or HTML file (for example, linkedin.htm). Lastly, you would open that file in a web browser to view the content.

Alternatively, you could paste that HTML into a third-party web page like codebeautify.org. Its web form will accept the HTML (1). Then, you click the Run button (2) and view the rendered HTML content on the right (3). Here, like in the local file discussed above, you can scroll up and down the LinkedIn content to see most all of the profile information!

Image from https://sec487.info/w-, October 3, 2019.

Google AdSense and Analytics

Both Google products, AdSense and Analytics, use unique numbers for each customer account

The same number is used across multiple sites the customer maintains

This common connection helps us find other sites that may help our investigation

The unique codes are found in the source code of the pages and look like:

- AdSense = pub-12345678
- Analytics = UA-87654321

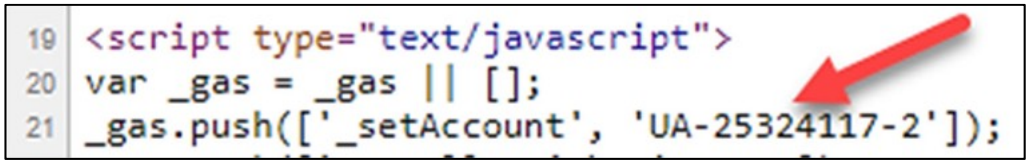
We can retrieve these manually or with tools and use sites to look up other sites with the same codes

Google AdSense and Analytics

Google has two services that are of interest to OSINT analysts: AdSense and Analytics. Both of these services provide their users with unique numbers that are to be embedded in the HTML source code of the pages that the customer owns and/or maintains. When delivered to the visitors of the web sites, JavaScript in the page sends the unique codes to Google's servers, either pulling ads, in the case of AdSense, or tracking that a user visited that page, with Analytics.

One AdSense or Analytics code can be used across multiple web sites and on a variety of web pages, with all the data flowing back to the Google customer's dashboards. As OSINT analysts, when we find these codes in web pages, we can sometimes look them up and find other web sites that also have those unique codes in them. When we are successful, we know that there is a relationship between both of those sites.

The codes can be found in the HTML source code of the pages that appear in a browser (image below from view-source:https://www.sans.org/) or are retrieved using other means, such as Wget, curl, and scripts. AdSense codes start with "pub-" and have six or more numbers after them (example: pub-12345678), whereas Analytics codes begin with "UA-" and then have numbers (example: UA-87654321).



Google Analytics Search Engines

http://spyonweb.com and http://sameid.net

Enter Google Analytics code in search field and results show other domains with that ID

Example on right is for **match.com**'s UA code

Google Analytics

UA-16351953 44 domains

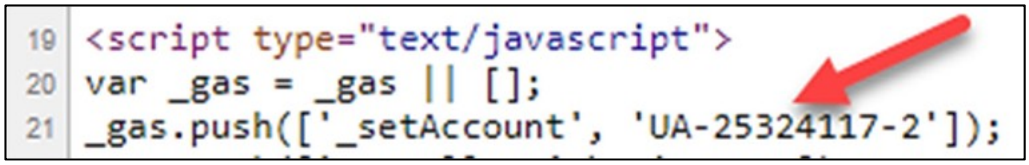
alternativematch.com	bachelorism.net
itallstartedwithalook.com	love.aol.jp
love.yam.com	majalahbisnisglobal.co
matchmyfriend.com	matchmyfriends.com
meetmoi.com	par-perfeito.com.br
singlesinamerica.com	singlesnet.com
udate.com	www.agenciadeencontros.com.br
www.agenciasdenamoro.com.br	www.alternativematch.com
www.bachelorism.org	www.conquistar-homem.com.br
www.dicas-de-paquera.com.br	www.e-llure.com
www.encontrarmulher.com.br	www.kiss.com

Google Analytics Search Engine

There are several places on the internet that collect the Google Analytics codes and store which web sites they were used on. http://spyonweb.com and http://sameid.net/ are two sites that are useful for these lookups. Enter the Google Analytics code harvested from the HTML source of a page, and other sites that display that same UA code will be shown. Above, we ran the UA code for Match.com (UA-16351953) and found 44 other domains that use the same code.

For more information on these techniques, or to see them in action, you may wish to watch the YouTube video for The OSINT Curious Project's 10-Minute Tip at https://sec487.info/qh.

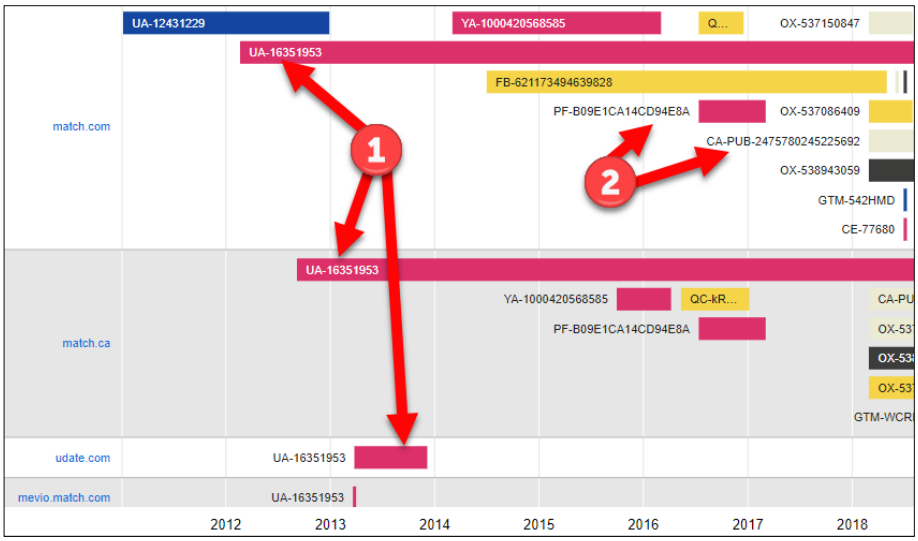
Image from https://sec487.info/ka, October 2, 2018.



Builtwith.com's Relationships

With builtwith.com we can examine trackers across time (1)

We can also see non-Google analytics codes (2) in sites



Builtwith.com's Relationships

Revisiting the builtwith.com web site, we can use its "Relationship Profile" feature to examine the tracking codes, analytics, and advertising bugs that it has found in related web sites across time. The image above highlights (arrow 1) a common Google Analytics "UA" code used across multiple sites and additional codes the site discovered (arrow 2). Beyond this information, builtwith.com also shows common IP addresses for these hosts. So if a domain resolves to an IP address that other domains resolved to, there is probably some kind of interesting relationship that may need to be investigated.

Image from https://sec487.info/qi, August 23, 2019.

Encryption Certificates

Web sites using HTTPS present TLS or SSL certificates to browsers to create the encrypted connections

These certificates can contain sensitive or private data we can harvest

When examining a web site, see if it has an HTTPS certificate

Information that can be obtained in a TLS or SSL certificate include:

- Domains
- IP addresses
- Emails
- Physical locations

Encryption Certificates

With more of the internet moving to encrypted HTTPS traffic, we may be able to harvest useful OSINT data from the certificates that make those connections encrypted. The process for configuring a server to use HTTPS to secure traffic includes the system owner/admin creating a Certificate Signing Request (CSR)[1]. The data they submit in the CSR can contain private data. The CSR is used to generate the certificate that the governing Certificate Authority signs and returns to the system owner to install on their server. Since the CSR might contain private data, the certificate will also contain that data. This is where we, as OSINT analysts, can benefit.

Inside the HTTPS certificate we can find data about the self-reported location, IP address, email address, and domains of the servers that the certificate may be installed into. Sometimes these certs have IP addresses from the internal (not internet-facing) networks and internal domain names too. All depends on how the submitter filled out the CSR.

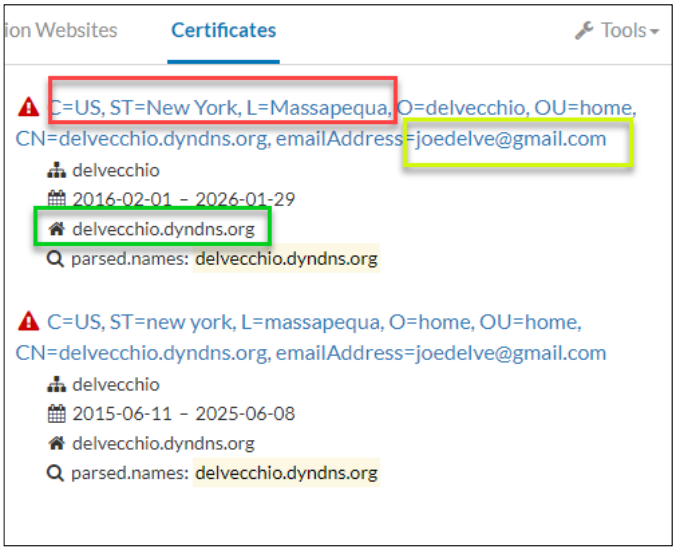
Reference:

[1] https://sec487.info/9t

Scenario: Suspicious Network Traffic

A practical application of this technique might come if your organization saw suspicious network traffic to the delvecchio.dyndns.org site

Searching for this site on censys.io yielded the certificate contents on the right



Scenario: Suspicious Network Traffic

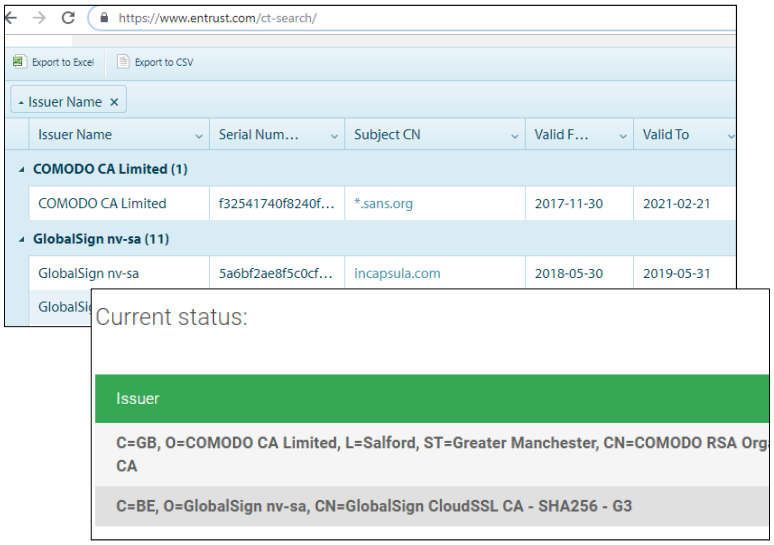
The OSINT we perform can be in support of many customers, one of which may be a Cyber Incident Response Team (CIRT). Let us see a practical application of evaluating the TLS/SSL encryption certificates in this example.

The CIRT submitted a request to your team to perform OSINT on the delvecchio.dyndns.org domain, as they saw suspicious network traffic coming from it in their logs. They want to know an address and phone number associated with it. You take the domain and query Censys.io (https://sec487.info/9r), pulling up two results, as shown in the slide. The certificate contents show the person who created it identified the location as Massapequa, New York, US and with an email address of joedelve@gmail.com. We can take that email and the location and use people search engines (https://sec487.info/9s) to discover more about the owner of the certificate (shown below). This data should help the CIRT find a person to help explain the suspicious traffic.

HTTPS Certificate Transparency

Certificate transparency sites that sometimes have different data and interfaces

- entrust.com
- transparencyreport.google.com
- crt.sh



HTTPS Certificate Transparency

There are web sites aside from Censys.io that allow you to search within SSL/TLS HTTPS certificate data. Each has roughly similar data but may display it differently or show expired and current certs. We can use these sites to understand the domains and IP addresses used by our targets, see where their systems are, and learn about other possible related domains that our targets manage or use.

Images from https://sec487.info/os and https://sec487.info/ot, January 5, 2019.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Harvesting Web Data" and "Web Analytics"

This page intentionally left blank.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

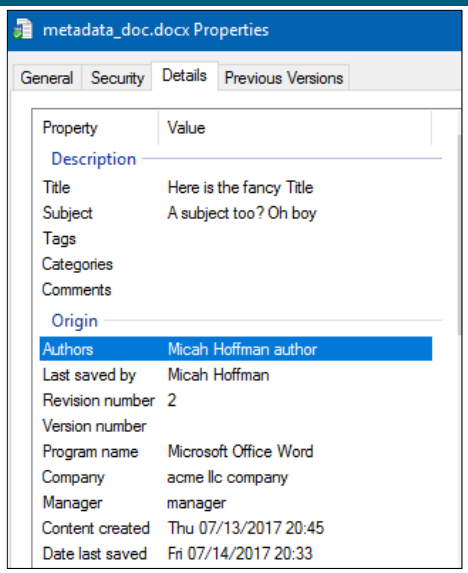
1. Data Analysis Challenges
2. Harvesting Web Data
3. **File Metadata Analysis**
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.

What Is Metadata?

Many of the files we use in our home and work lives contain extra information: metadata

This content is usually stored in attributes within the file but not normally displayed in the regular content



What Is Metadata?

Many of the documents we use in our work or daily lives contain extra information about the files. In Microsoft Word documents (as shown in the picture above), that additional information may be the title and subject of the document as well as the author and company's names. In photos, there may be positioning data and content about the shutter speed and lighting conditions. While not truly hidden, metadata is not usually shown to users when they use the document in the usual fashion. Open up the metadata_doc.dox file from the above screenshot in Microsoft Word and you will not see the title or subject in the main body of the document.

Microsoft has a good web page that describes the metadata you can find within a Word document.[1] The content includes comments and version information, document properties, headers/footers, hidden text, document server properties, and custom XML data.

For metadata within images, you may find the post at https://sec487.info/gk to be helpful as it describes the metadata tags that can be found in these files.

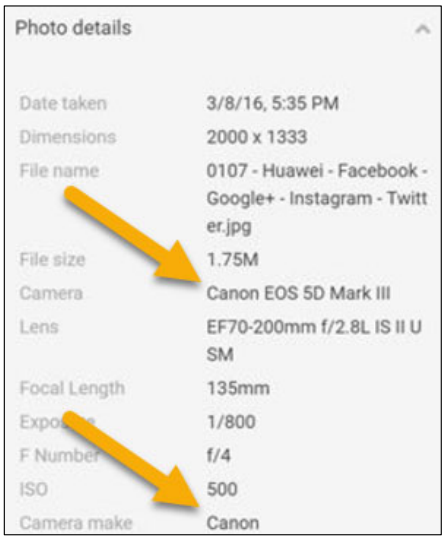
Once you know about metadata, it is trivial to retrieve it. We can accomplish this using standard methods within common operating systems or special tools (discussed later).

Reference:

[1] https://sec487.info/5s, August 16, 2017.

Huawei or Canon?

Huawei tried to pass off this picture as taken with a Huawei P9 phone, but metadata shows differently



Huawei or Canon?

Apple ran an advertising campaign that showcased pictures taken with their iPhones to show off their terrific cameras and image-processing abilities. In 2016, Huawei decided to do something similar with their P9 phone (https://sec487.info/m7) but, instead of using their phone to take the pictures, they used a high-quality, professional camera (Canon EOS 5D Mark III, https://sec487.info/m8) with an expensive lens.

When Huawei's marketing team posted the picture to the Google+ social media network, they didn't know that Google+ allowed viewers of the photos to see its metadata. Viewers on the internet found out about this and called Huawei out on the deception.

Image from https://sec487.info/60, August 19, 2017.

Why Is Metadata Important for OSINT?

Metadata can contain sensitive content that users do not realize they are divulging

Users may forget to remove metadata from their files before publishing/sending them

It provides leads and can give insights into the OSINT target by showing us:

- user names
- GPS locations
- dates and times
- file locations
- timestamps (create, access, modification)

Why Is Metadata Important for OSINT?

While our OSINT targets may be careful about the contents of an image file (what they took a photo of) or document, such as a PDF or Excel Spreadsheet, the metadata that accompanies the file may contain sensitive data that the user has forgotten to remove.

What metadata is stored in a file depends upon the document type and what application or hardware created it. In word processing documents we may expect to see parameters such as title, author, and subject, whereas in media files we may view the bit rates for audio tracks, resolutions for movies, and apertures of lenses used to capture images.

Common metadata stored across different file types include creation, access, and modification timestamps, user names, file system paths, and names and versions of the software used to create or modify the file.

What Files Have Metadata?

People mainly think of metadata associated with images

But there are over 170 different file types that may contain metadata

Some websites might remove or strip metadata from files when they are uploaded

Examples include:

- Office documents, spreadsheets, and slides
- Movies (MP4, MOV)
- Audio files (MP3, WAV)
- HTML files
- Images (JPG, PNG, TIFF)
- PDFs
- Compressed archives (RAR, ZIP)

What Files Have Metadata?

There are over 170 types of files that may contain metadata. Everything from common "office" files (documents, spreadsheets, and slide decks) to images, videos, and audio files can have some types of metadata inside them. When you find a file posted online, consider checking it for metadata.

While many files CAN have metadata, oftentimes web sites will remove that metadata before posting uploaded documents. So when you try to download a file from a web site, the metadata that was in it may have been removed.

How Do We Read Metadata?

The Exiftool website shows that over 170 different file types have some metadata

When you download or find a file, check it for metadata

Viewing the metadata in a file can be accomplished using several techniques:

- Command-line tools
- Operating system tools
- Specialized tools
- Web sites

How Do We Read Metadata?

One of the metadata extraction tools we will encounter, the Exiftool, posts over 170 different file formats[1] that contain metadata. The lesson here is, "When you find a file, examine it for metadata." You never know what you will find.

Accessing the metadata can be accomplished using four main methods:

- Using scripts and binaries that we run from a command prompt or terminal window.
- Using the built-in viewers and dialogs in macOS and Windows operating systems.
- Specialized tools can help us gather this data in bulk, especially when analyzing many files.
- Finally, there are some web sites that will display metadata if we upload a file.

Another resource for accessing file metadata is the ForensicsWiki,[2] which shows how to extract metadata from a wide variety of sources using web sites and local programs.

Justin Seitz (@jms_dot_py) wrote a great blog entry on scraping metadata from files archived on Archive.org using a Python script (https://sec487.info/5z). Python code is included!

References:

[1] https://sec487.info/5o, July 14, 2017.

[2] https://sec487.info/5t

Example Image with Metadata (1)

This image has metadata that we can extract

We can use tools or web sites to view the data

Image metadata might have camera information, dates and times, and GPS locations

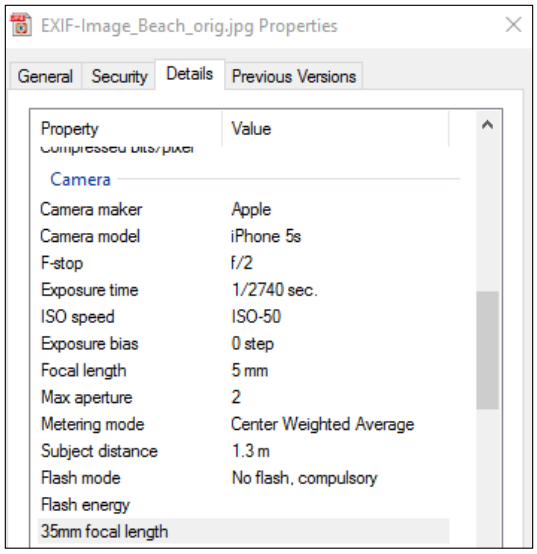
Example Image with Metadata (1)

As OSINT analysts, we deal with images and can use the metadata stored in a file to understand what device took the picture, when it was taken, and where (provided that the data has not been tampered with). Above is a typical beach picture of sun, surf, and sky. A few rocks can be seen in the water as well. Let us look at the metadata in the picture.

Windows Properties

In modern versions of Microsoft Windows, metadata can be viewed and edited in the properties of the file within Windows Explorer

Right-click a file, choose the Details tab, the data appears and can be changed



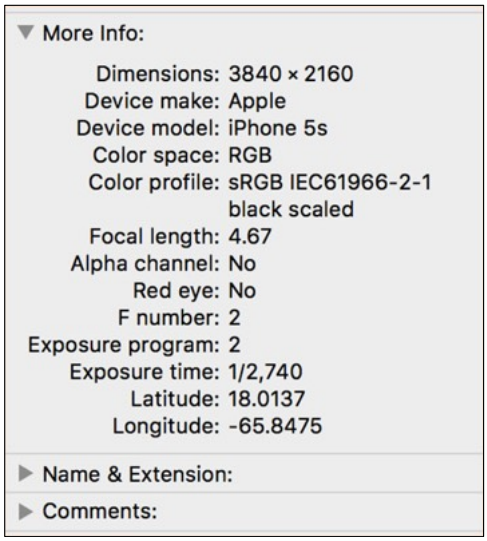
Windows Properties

Modern versions of Microsoft Windows make viewing and editing metadata extremely simple. From the file system, right-click on a file that you wish to see or edit the metadata of and then select the "Details" tab. The metadata will show up as name and parameter value pairs in the window (as shown above in the picture). If you wish to edit the data, simply double-click the value you wish to change, enter in the new content, and then click the "OK" button to save the data with the file.

macOS Properties

macOS allows users to read metadata using the Get Info feature in the Finder

Select the file you wish to see the data in, select Get Info, and then expand the "More Info" section to view some data



macOS Properties

Natively, macOS allows viewing of some metadata from the user's desktop by right-clicking on a file and choosing the "Get Info" option. In the box that pops up, expand the "More Info" section to see some data. This is a read-only view of the content, and only some of the metadata present is displayed. Using Brew or Ports, users can add the Exiftool to macOS and leverage its abilities to perform other metadata manipulations.

To view more metadata from pictures, the Preview application allows users to use the Tools -> Show Inspector menu options to open the Inspector window. Clicking on the "More Info" tab that has an "i" in a circle, displays all the EXIF data in the picture as shown below. Choosing a different sub-tab (General, Exif, GPS, JFIF or TIFF) will display metadata but not will not allow the user to edit that data.

Web Site Metadata Analysis

One of the most accessible methods of viewing metadata is via web site viewers

Many of these sites only process images

OPSEC WARNING: Sharing files and metadata from your customer with third-party sites may not be desirable

There are a variety of metadata sites to choose from, and we like the following:

- http://www.verexif.com
- https://www.thexifer.net

Web Site Metadata Analysis

If downloading a special tool to view the metadata in a file is distasteful to you, then you will be happy to know that there are a variety of web applications that will assist you. Typically, a user will upload a document or image to one of these web sites, and then the site extracts and displays the metadata on screen for the user. While web sites make viewing metadata easy, most of them are configured to only process image files.

OPSEC WARNING: Prior to uploading any assessment pictures or documents to a third-party web site, evaluate if your customer and employer will allow this. Many times, these stakeholders will want you to perform local analysis of the files so that others do not have that data.

Two of our favorite web applications for processing image metadata are verexif.com and thexifer.net. Each of these sites will display metadata extracted from an image and also allow you to remove it (verexif.com) or tamper with it (thexifer.net). Performing a Google search for online "metadata viewer" yielded over 30,200 results (August 16, 2017), so you have many choices.

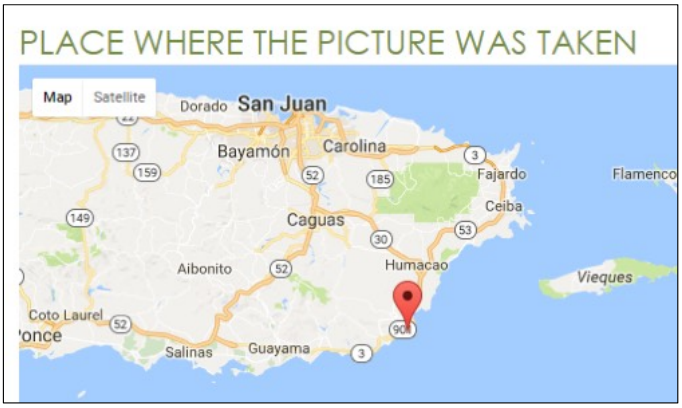
Example Image with Metadata (2)

EXIF DATA

Camera make :	Apple
Camera model :	iPhone 5s
Date/Time :	2017/04/14 13:50:26

GPS Latitude :	N 18° 0' 49.39"
GPS Longitude :	W 65° 50' 49.91"

http://www.verexif.com



Example Image with Metadata (2)

Using the http://www.verexif.com site to analyze the metadata of the previous image, we can extract some vital data.

- Camera make: Apple
- Camera model: iPhone 5s
- Date and time the picture was taken
- GPS location where the picture was taken

This web site even provides a map where it has plotted the GPS location reported in the image's metadata. We see that this location is found on the island of Puerto Rico.

Does this metadata matter? Depends on the investigation and what the data shows and does not show. We will examine several methods of extracting this information from files.

Command Line - Exiftool

Exiftool is a command-line tool that works on Windows, macOS, and Linux

Can read, edit, and remove metadata from files

Online documentation and help (man) files show many switches and flags that may be useful

exiftool EXIF-Image_Beach_orig.jpg

```
$ exiftool EXIF-Image_Beach_orig.jpg
ExifTool Version Number         : 10.80
File Name                       : EXIF-Image_Beach_orig.jpg
Directory                       : .
File Size                       : 1396 kB
File Modification Date/Time     : 2017:07:13 20:58:28-04:00
File Access Date/Time           : 2018:06:14 18:49:41-04:00
File Inode Change Date/Time     : 2018:06:14 18:49:41-04:00
File Permissions                : rwxrwxrwx
File Type                       : JPEG
File Type Extension             : jpg
MIME Type                       : image/jpeg
Exif Byte Order                 : Big-endian (Motorola, MM)
Make                            : Apple
Camera Model Name               : iPhone 5s
Orientation                     : Horizontal (normal)
X Resolution                    : 72
Y Resolution                    : 72
Resolution Unit                 : inches
Software                        : HDR+ 1.0.147548766z
Modify Date                     : 2017:04:14 13:50:26
Y Cb Cr Positioning             : Centered
Exposure Time                   : 1/2740
F-Number                        : 2.0
```


Command Line - Exiftool

The Exiftool (https://sec487.info/5o) is the go-to tool for most command-line-savvy people. With Linux, Windows, and macOS packages, it is a powerful and flexible cross-platform application. Exiftool allows the user to read, write, and remove metadata from files.

Documentation on how to use the tool can be found at https://sec487.info/5p. If you are using a macOS or Linux system, consider using the "man exiftool" command inside a terminal window to get more assistance on exiftool's features.

In the picture above, we tell the Exiftool to read the metadata from the EXIF-Image_Beach_orig.jpg picture. The full output of the tool contained much more than is displayed in the picture above. Viewing the above information, we can see that the picture was taken on April 14, 2017 at 1:50pm (13:50:26) from an Apple iPhone 5s.

If we wished to modify or delete the metadata, instructions with examples can be found at https://sec487.info/5q.

FOCA (Fingerprinting Organizations with Collected Archives)

Older tool from Eleven Paths but still very relevant and useful

Windows-only application that:

- Will search Google for documents about a certain domain
- Download those documents
- Extract and analyze the metadata from those files

Finds, downloads, and processes most major file types, including "office" docs and image files

Can save analysts time when analyzing many files from a single domain

FOCA can also perform web attacks against target sites, so be careful how you configure it

FOCA (Fingerprinting Organizations with Collected Archives)

FOCA is a tool that has been well known for metadata analysis of files for many years. While FOCA is not actively developed by its owner, Eleven Paths, it has many features that work well, even with some of the newer file formats we see on the internet. The downloadable, Windows-only application (https://www.elevenpaths.com/labstools/foca/index.html, https://sec487.info/5u) is easy to install and run.

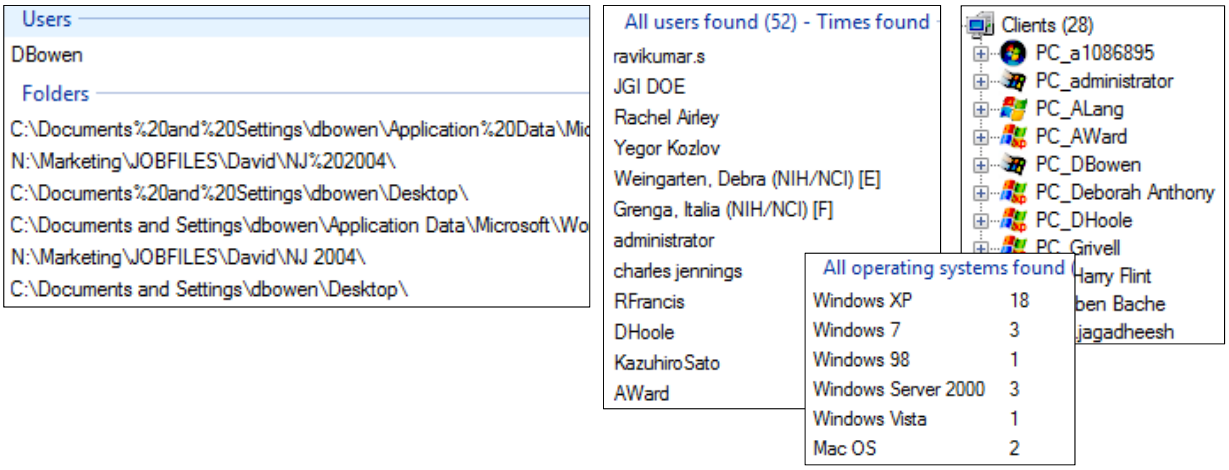
The process that FOCA uses in its work can be summarized as follows:

- The user configures the tool and enters a domain of interest.
- FOCA searches the internet and any web sites that it finds for documents of a certain type (specified by the user before running the tool).
- Once FOCA finds documents, the user requests to download them locally to their computer.
- The user instructs FOCA to extract and analyze the metadata in the files.

FOCA has features that can discover and exploit web site vulnerabilities if the user wishes to. Since we, as OSINT analysts, are interested in metadata, these features may not be as useful.

FOCA Output

Analyzed public documents on the nature.com web site in August 2017



FOCA Output

Using the FOCA client, we examined a mix of over 50 Microsoft PowerPoint, Excel, and Word documents found on the nature.com web site in August 2017. We ran FOCA, downloaded the documents, extracted the metadata, and then analyzed it. The slide above shows some of the content that was discovered hidden in the metadata.

Some interesting items found in the extracted data include:

- Some of the documents were written on Windows 98 and Windows XP systems.
- Authors of the documents created some on mapped drives such as the "N" drive above.
- We found over 50 user names that showed full first and last names or the naming scheme of the user (first initial + last name).
- The names of over 25 computers (shown as "clients" above) were discovered.

FOCA is faster than a human would be at what it does and can accelerate your OSINT investigation.

Tampered Metadata

Original

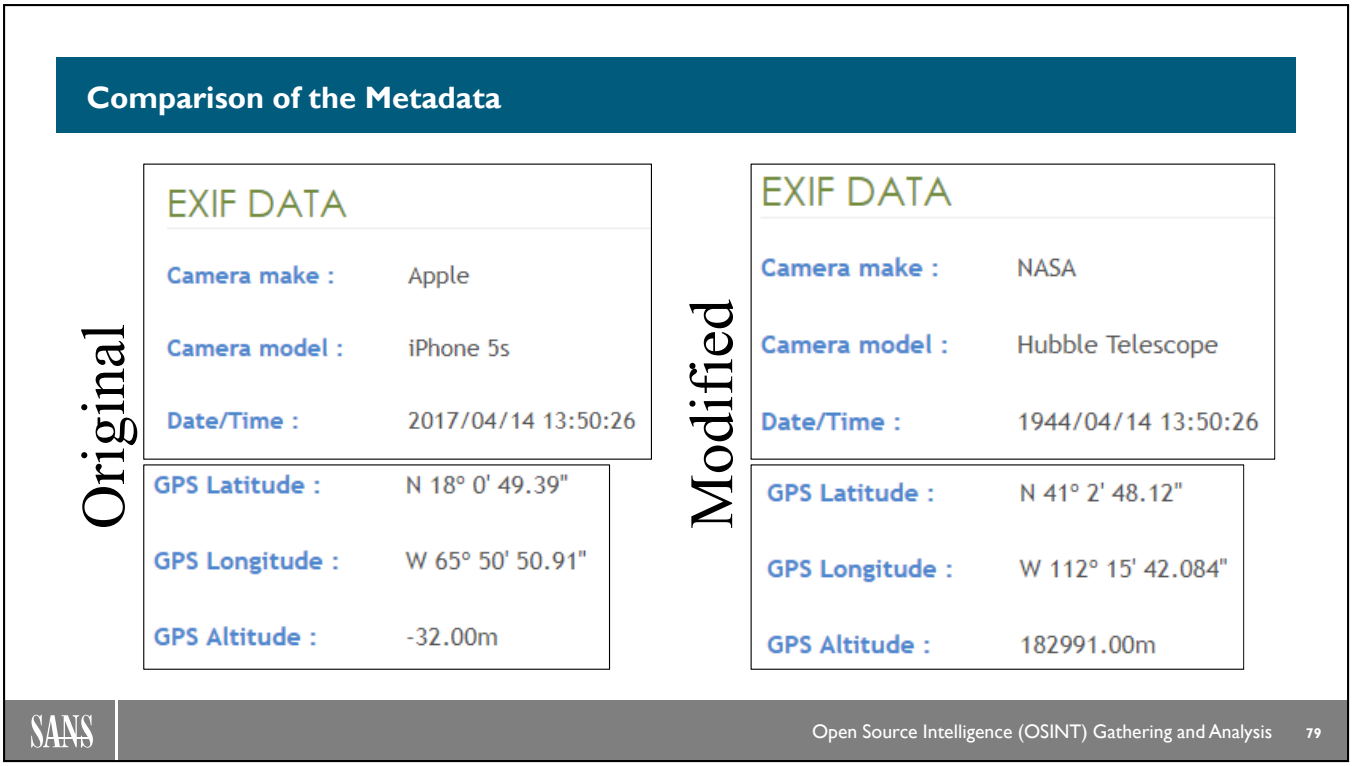
Modified

Tampered Metadata

We used the above images as examples of files with metadata. As discussed, there are many methods to extract the GPS metadata from the images. The next question you might ask yourself is, "Are the GPS coordinates accurate?" We can tell if the metadata has been tampered with if it contains conflicting or wrong data. Let us verify the GPS coordinates the picture had and see if that data was falsified by taking note of the physical features in the picture and trying to match them to features at that GPS location in an online mapping application.

Josh Huff (@baywolf88) wrote more about this process in his blog post "GOOGLE MAPS... MORE OSINT THAN YOU THINK" (https://sec487.info/5r). He detailed how he found an intersection where there was police activity from OSINT data.

The images in the above picture appear to be identical and, in fact, they are the exact same picture. The image on the left has the original metadata, and the one on the right has had its data changed. Let's examine how we can verify the metadata.



Comparison of the Metadata

We used the http://www.verexif.com web site to analyze the metadata in the files and extracted the content above. Clearly one is plausible and one is less plausible. In your OSINT assessments, the tampered data may not be as silly as what we did in the picture above (it is doubtful that the United States National Aeronautics and Space Administration would allow its Hubble space telescope to be used to take this Earth-bound image).

While the camera make and model appear altered, what about the date? Could that be factual? It is possible that someone digitized an old photo and, through post-processing software, added the correct date to the new digital image's metadata.

What about the GPS coordinates? Those are different in the two images. We would need to perform additional work to discover if either of them is correct.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Metadata Analysis"

This page intentionally left blank.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. **OSINT Frameworks**
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.

OSINT Frameworks Working for You

What if there were collections of tools and web sites useful for OSINT investigations?

What if there were sites that helped you understand how to transform one type of data (e.g., email) into another type of data (e.g., social media profiles)?

There are, and they are called frameworks.

OSINT Frameworks Working for You

OSINT investigators across the world use many of the same web sites in their work. With the vast number of web applications, data stores, and "dark places" on the internet, it can be challenging to remember all of the most useful locations where we can find information. This is where frameworks come into play; they act as a giant, well-organized bookmark page for OSINT links and tools. They also can help us understand what sites can help transform one type of data into another. Say, for example, that you found an email address for your target. Leveraging one of these frameworks, you can find links and tools to help you find social media profiles, phone numbers, and business listings.

In the coming slides, we will examine a couple helpful frameworks and some other web sites that we can use for OSINT.

OSINT Framework

Justin Nordine (@jnordine) created an open source, visually appealing method of viewing OSINT URLs simply called the “OSINT Framework"

Rendered version is at https://osintframework.com/

Source code/project at https://sec487.info/15

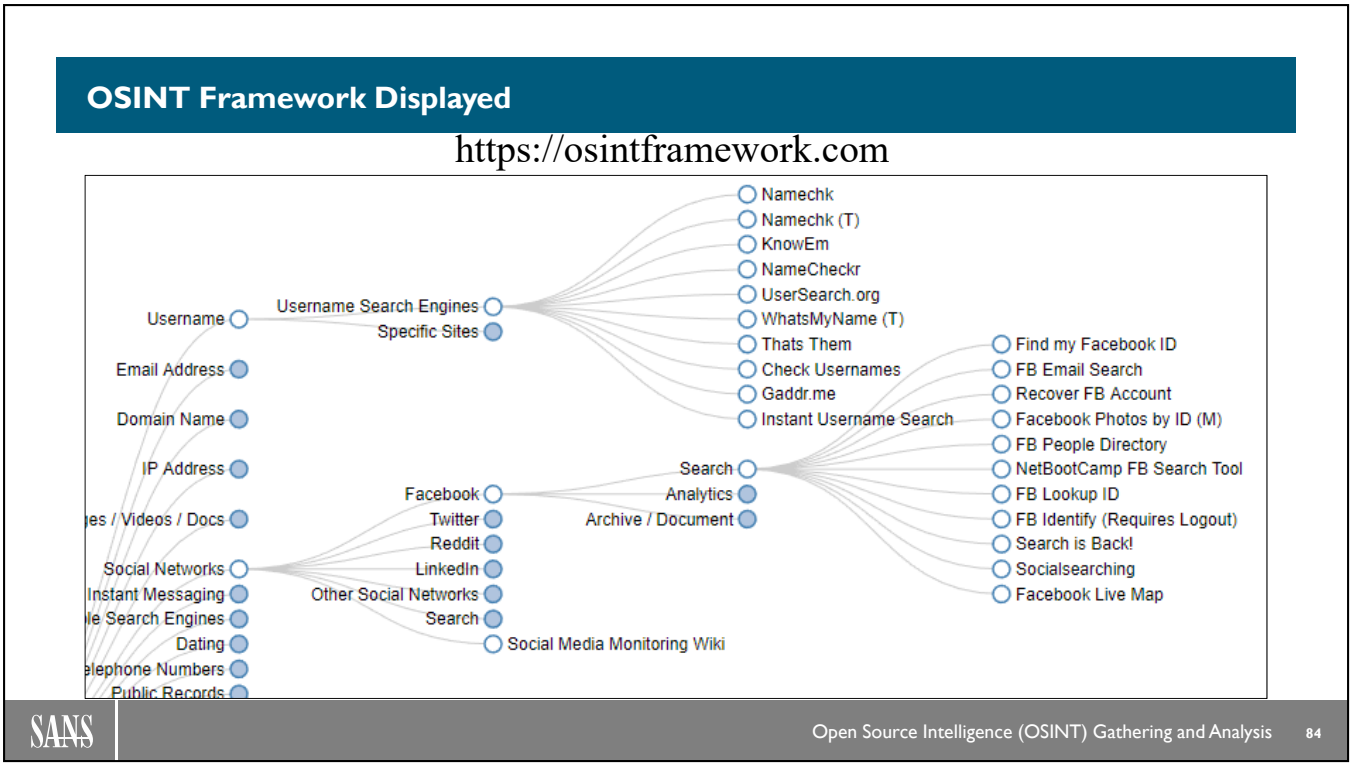
Can be used offline or customized to your team's needs.

OSINT Framework

Justin Nordine (@jnordine) took the core content from the PTES, combined it with some other resources, updated the links, added some flashy JavaScript and made it an excellent, up-to-date resource. The main site is online at https://osintframework.com/.

Justin has made this an open source project and hosts the source code for the site on GitHub at https://sec487.info/15. Having access to the source code can be helpful for several reasons:

1. You can download the entire content of the framework and use it locally or on a closed network.
2. You can edit the content to add your own links.
3. You can contribute to the project by editing the source code and submitting it to Justin for inclusion. ← GREAT opportunity to give back to the community!



OSINT Framework Displayed

The rendered version of the framework is a useful tool to remember different web sites from which OSINT data can be retrieved. Using the D3js Data-Driven Documents (https://d3js.org/) JavaScript code, the OSINT Framework visually connects each of its resources in a parent-child relationship. This can also be thought of as a directory-file relationship since a user clicks a parent (or directory) and it expands to show its child nodes, some of which may be directory-like nodes and some will be links to other web sites. In the slide above, the parent-child relationship is displayed for the "Username" branch of the project.

The use case for the OSINT Framework is that if you are looking for or have certain data and need places to look up what to do with it or places to transform it, the OSINT Framework has links to applications and other sites you can use. With its easy-to-use point-and-click interface and the categories, it makes finding new places to use for your OSINT investigations trivial.

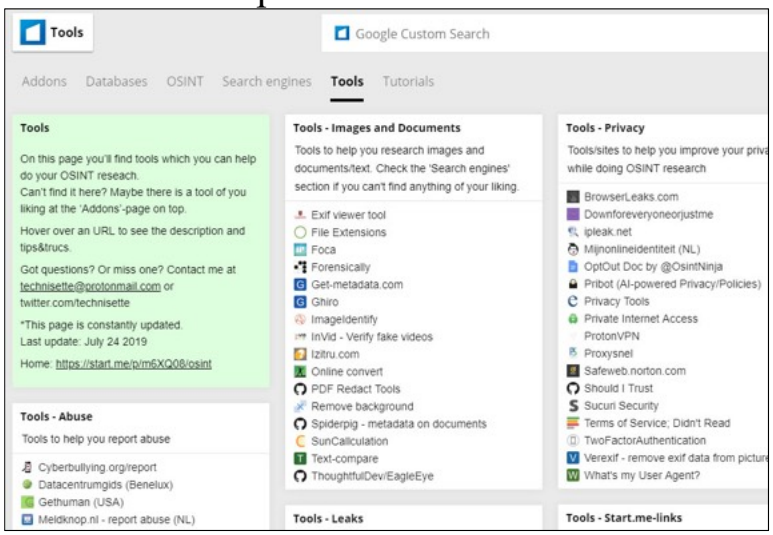
Image from https://osintframework.com, August 23, 2019.

technisette.com (start.me page)

The Netherlands-based OSINT analyst "technisette" curates an amazing OSINT resource page

https://technisette.com redirects to the start.me page with tools, techniques, and other resources

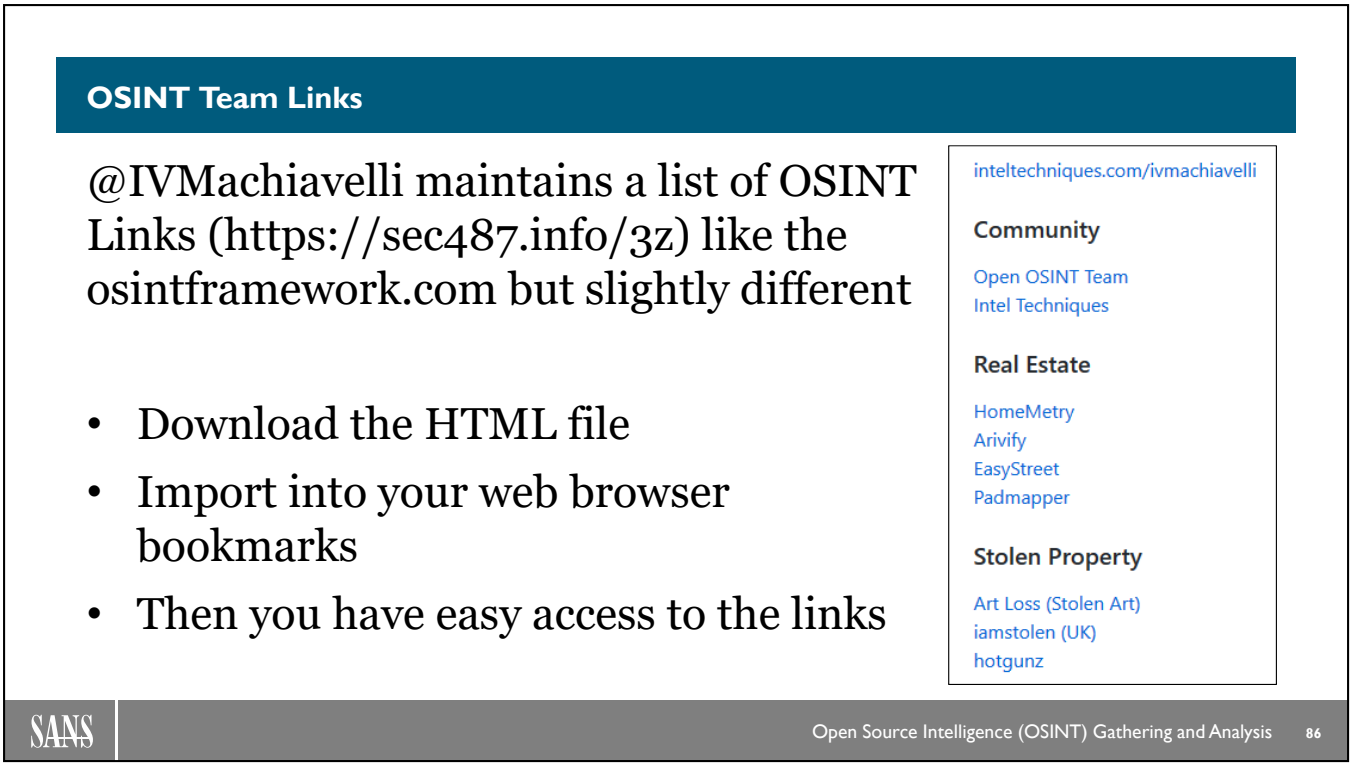
http://technisette.com



technisette.com (start.me page)

A person in the Netherlands who uses the name "technisette" (Twitter: @technisette) created a domain with her name (https://technisette.com) that redirects user web browsers to an amazing OSINT resource. She has curated a start.me page with OSINT talks, tools, techniques, and tutorials. Some of these resources overlap with others we have seen, but the visual layout of her links and the way they are grouped make them easy to find and use.

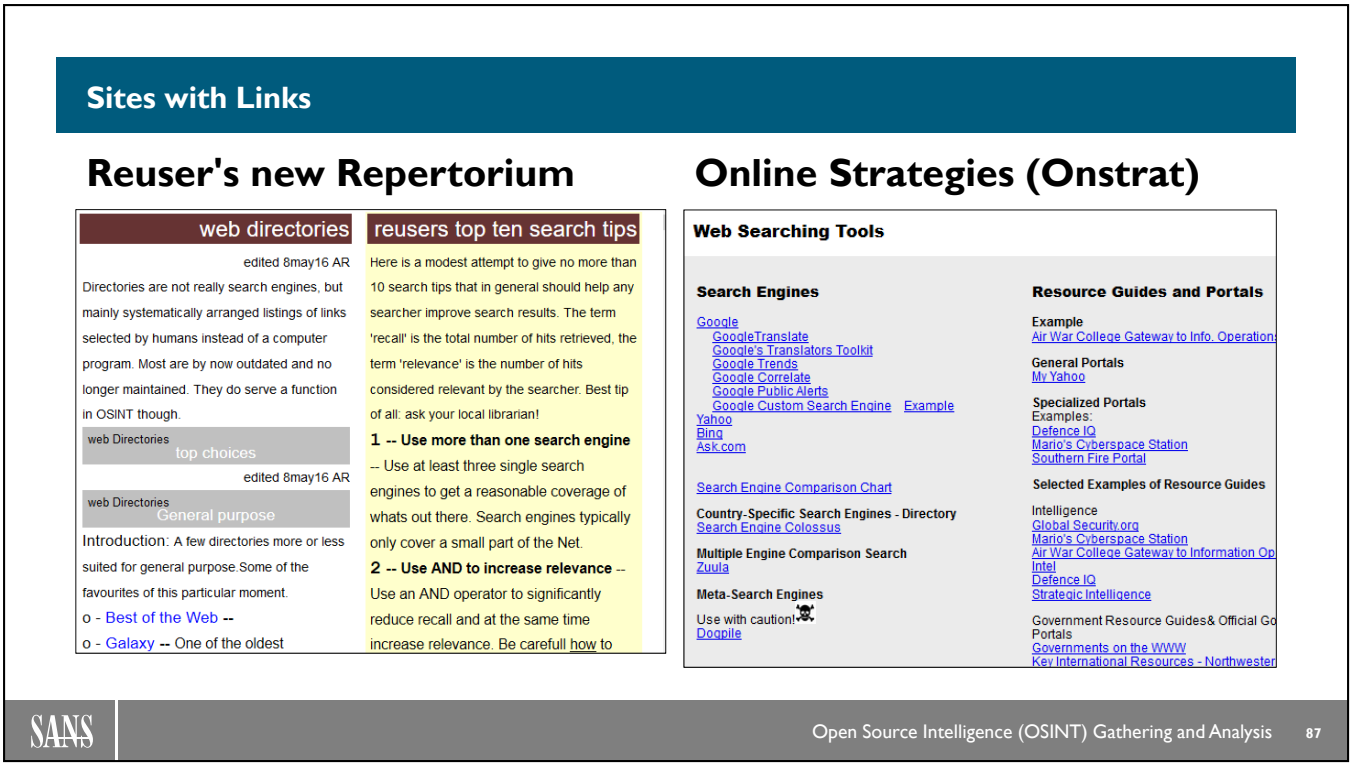
Image from https://sec487.info/qf, August 23, 2019.



OSINT Team Links

A web user by the user name "IVMachiavelli" has created several useful OSINT resources. You may remember this name as the creator of the Open OSINT Slack group mentioned earlier in the course. They also have a GitHub project that compiles and organizes OSINT URLs based upon the types of data you wish to find. It is well curated and actively maintained and available at https://sec487.info/3z.

To use the resources, download the HTML document and either import it into a web browser or open it in a web browser to access the links.



Sites with Links

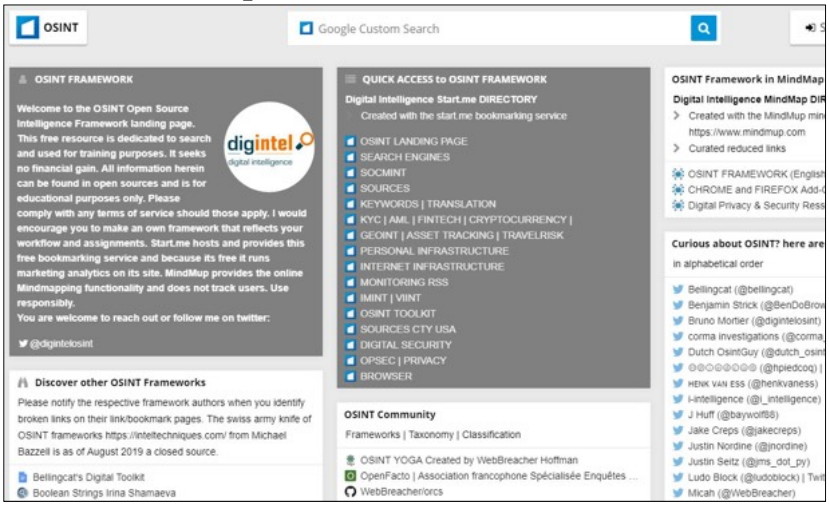
Next, we have the Reuser's new Repertorium (https://sec487.info/kc) and the Online Strategies (https://sec487.info/kb) web sites. To say that they also have links to useful sites would be an understatement. The Reuser's site even has some tips for OSINT investigations!

Images from https://sec487.info/kb and https://sec487.info/kc, December 12, 2018.

Bruno Mortier's OSINT Page

Bruno Mortier (@digintelosint) has a start.me page that lists other's OSINT framework and bookmark pages

http://osintframework.de



Bruno Mortier's OSINT Page

Because there are so many OSINT bookmark and framework sites, we need a good way to keep track of them. Bruno Mortier (@digintelosint) created an excellent start.me resource at https://sec487.info/ia that does just that. It shows other's sites and projects.

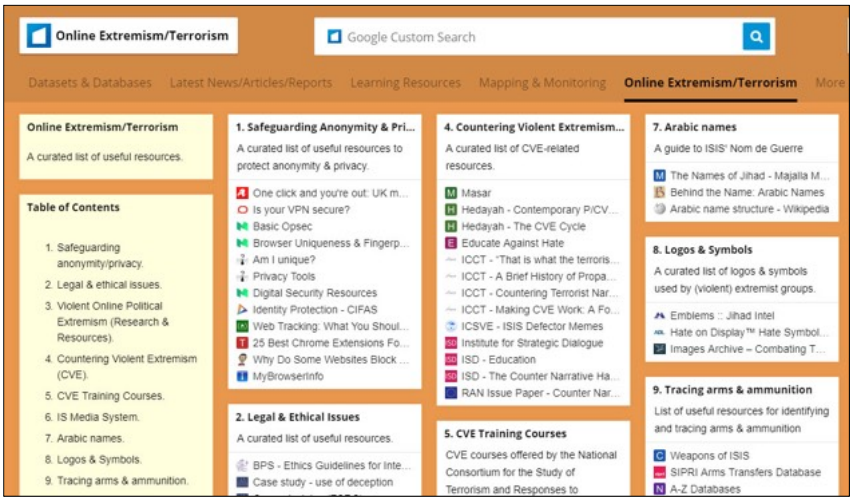
Image from https://sec487.info/ia, August 23, 2019.

Lorand Bodo's Terrorism & Radicalisation Research Dashboard

Looking for international terrorists and the systems and tools they use?

Try this resource to understand how and where they work

https://sec487.info/it



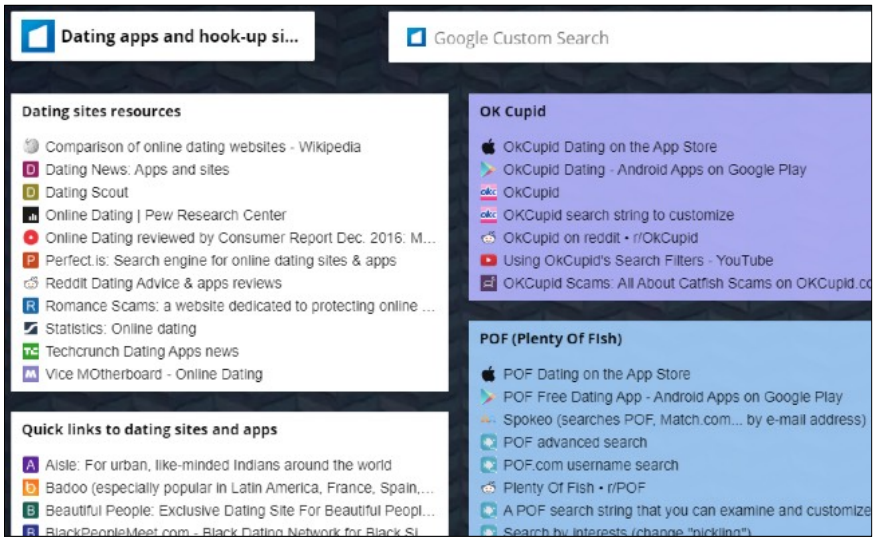
Lorand Bodo's Terrorism & Radicalisation Research Dashboard

With categories like FOSINT (Financial OSINT) and Crime-Terror Nexus, the Lorand Bodo's Terrorism & Radicalisation Research Dashboard (https://sec487.info/it) could be your go-to resource for understanding terrorists and hate groups.

Image from https://sec487.info/it, August 23, 2019.

Excellent Dating Resource

Emmanuelle Welch created the amazing start.me page cataloging a large number of dating and sex sites



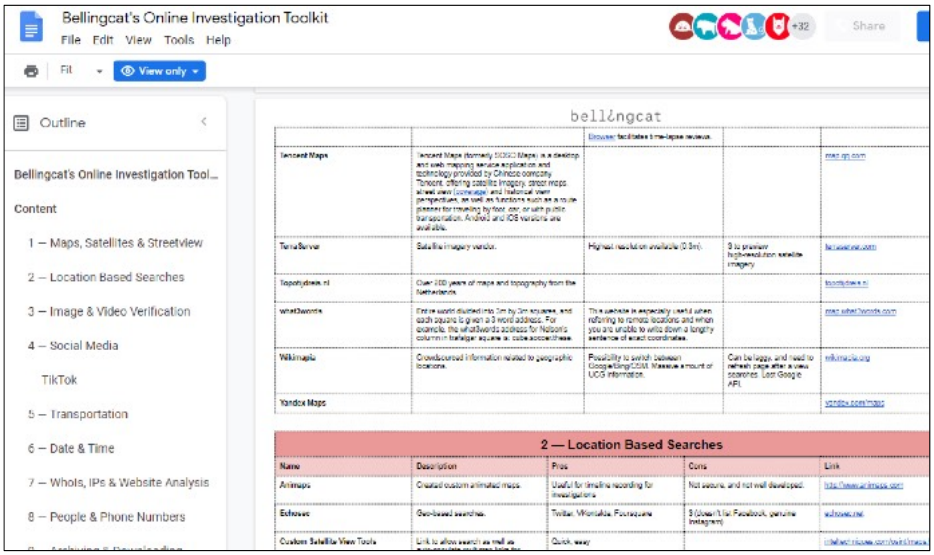
Excellent Dating Resource

Emmanuelle Welch (@frenchPI) created and maintains an excellent resource (https://sec487.info/ca) for OSINT analysts that need to venture into the world of dating and sex sites. She has links and tips there for a variety of sites.

Image from https://sec487.info/ca, September 6, 2019.

Bellingcat's Online Investigation Toolkit

Bellingcat maintains a Google document with well-categorized links for international OSINT



Bellingcat's Online Investigation Toolkit

We mentioned the Bellingcat organization already in this course. They are an excellent source for controversial, in-depth, OSINT-related journalism. Their organization performs OSINT on some of the most heinous crimes and hardened criminals the world has. From investigating the MH17 airplane disappearance to researching Neo-Nazi groups, Bellingcat tackles the challenging OSINT tasks.

They posted their online resource suggestions in a Google document at https://sec487.info/Bellingcat**.**

Image from https://sec487.info/bellingcat, September 17, 2019.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. **Basic Data: Addresses and Phone Numbers**
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.

Basic Public Data: What Is It?

Data that is not private and is in the public arena

Can usually be found on the internet

You may not be able to stop them from being published

Examples include:

- Home and work street address
- Phone numbers for home, work, and cell phones

Basic Public Data: What Is It?

Years ago, people in many countries received a telephone book on their doorstep once a year. This book contained the names, addresses and phone numbers of people in their region that had telephones. Much of this same information, and some additional pieces, can be thought of as basic public data.

This data may be included in a variety of online databases and search engines that can be easily queried. Much of it cannot be removed from the internet as it is in the public record. The sale value of a home, the court records from a divorce, and business registration data all are found on internet-facing web sites and in official documents that can be retrieved by visiting a courthouse.

Why Gather Street Addresses

Ties web data to physical locations

Ties people to property (other offices, vacation homes ...)

Allows us to perform other actions (social engineering, physical penetration testing, surveillance) against the target

Why Gather Street Addresses

Physical street addresses, whether they are for residences, businesses, or other locations, can be valuable pieces of data, depending upon what your overall OSINT goals are. They represent actual locations in our real world, and when we can tie an activity we learn about on the internet to the real world, it can open important avenues of exploration. Most physical addresses (for example, the Burger King restaurant near the Kremlin has an address of Manege Sq, 1, Moskva, Russia, 125009) are easily retrievable from internet sources and are deemed public data. With a target's or target company's physical address, we can prepare a variety of other reconnaissance activities, such as aerial surveillance or stakeouts, physical attacks (examples: kidnapping, property theft, etc.) and cyber attacks (examples: social engineering, physical penetration tests, and USB drive drops).

Physical addresses also tie people and entities to property. This can be useful in OSINT investigations for finding other homes or offices your targets may own or rent. Using these techniques, we can also discover people's neighbors, co-tenants, and relatives who may be living in the same home or near our target.

Image from https://sec487.info/eg, September 27, 2017.

Street Addresses Are Everywhere

We can gather street addresses from most places on the internet. Some favorites of OSINT researchers include:

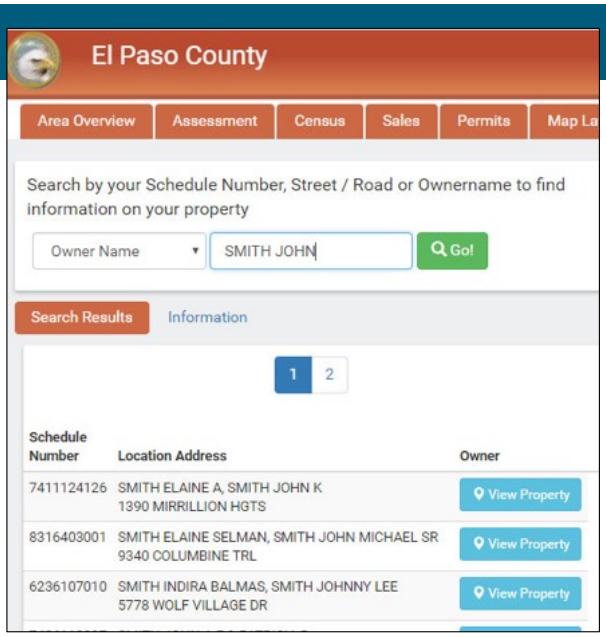
- Many localities have real estate taxes that may be publicly searchable. Using a target's name or company may show you these documents. Since these documents are official, we trust their accuracy more than other sites.
- People search engines such as thatsthem.com and whitepages.com will pull up addresses (and other data) when results for a target are shown.
- Resumes and business documents can contain addresses for personal and business locations.
- Company web sites frequently have an "about us" or "location" page that contains their physical address or addresses.
- Mapping web sites such as Google Maps and Bing Maps can be used to browse an area. When you find a place that you'd like the address of, click or right-click the place and choose to reveal the address.
- Real estate web sites (also terrific places to gather floor plans and demographic data)

Let's investigate a few of these places a bit closer.

Real Estate Tax Documents

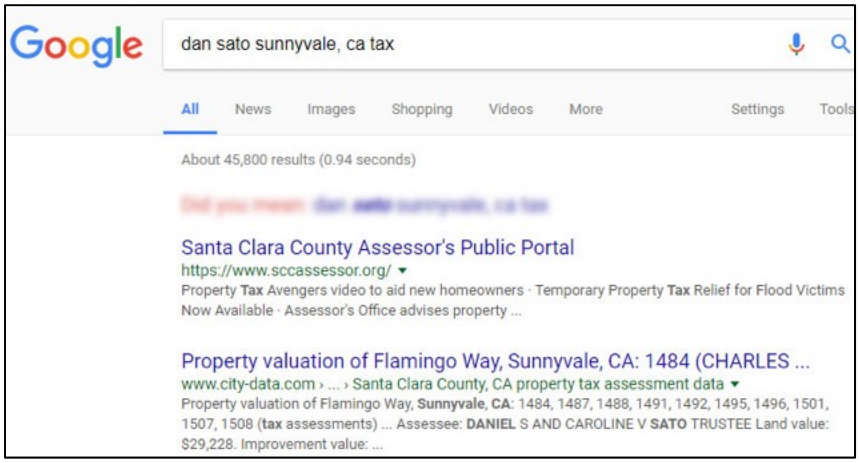
In the United States and a few other countries, property or real estate tax documents are posted to the internet

Some states, counties, and cities have search pages allowing you to enter a person's name, and some allow the data to be indexed by search engines



Real Estate Tax Documents

Within the United States, OSINT analysts can search and collect government data such as real estate taxes (also known as property taxes). Sometimes retrieving relevant tax data is as simple as performing a search engine query for your target's name and a potential geographic location (for example, John Smith in El Paso, Colorado, as shown in the picture above). We can use search engine directives such as `"TARGET NAME" LOCATION tax` (example: `"john smith" Detroit tax`) to retrieve possible indexed entries (such as the one below showing the county search web site and an entry for a person).



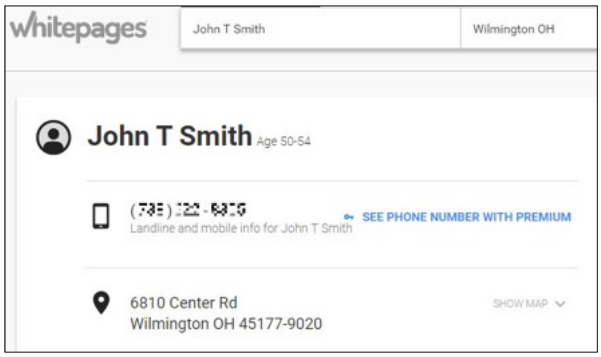
Images from https://sec487.info/eb and https://sec487.info/ec, July 1, 2017.

People Search Engines

People search sites for people inside the United States and Canada are common

Enter the name of your target into the site's search and then narrow down the results using filters

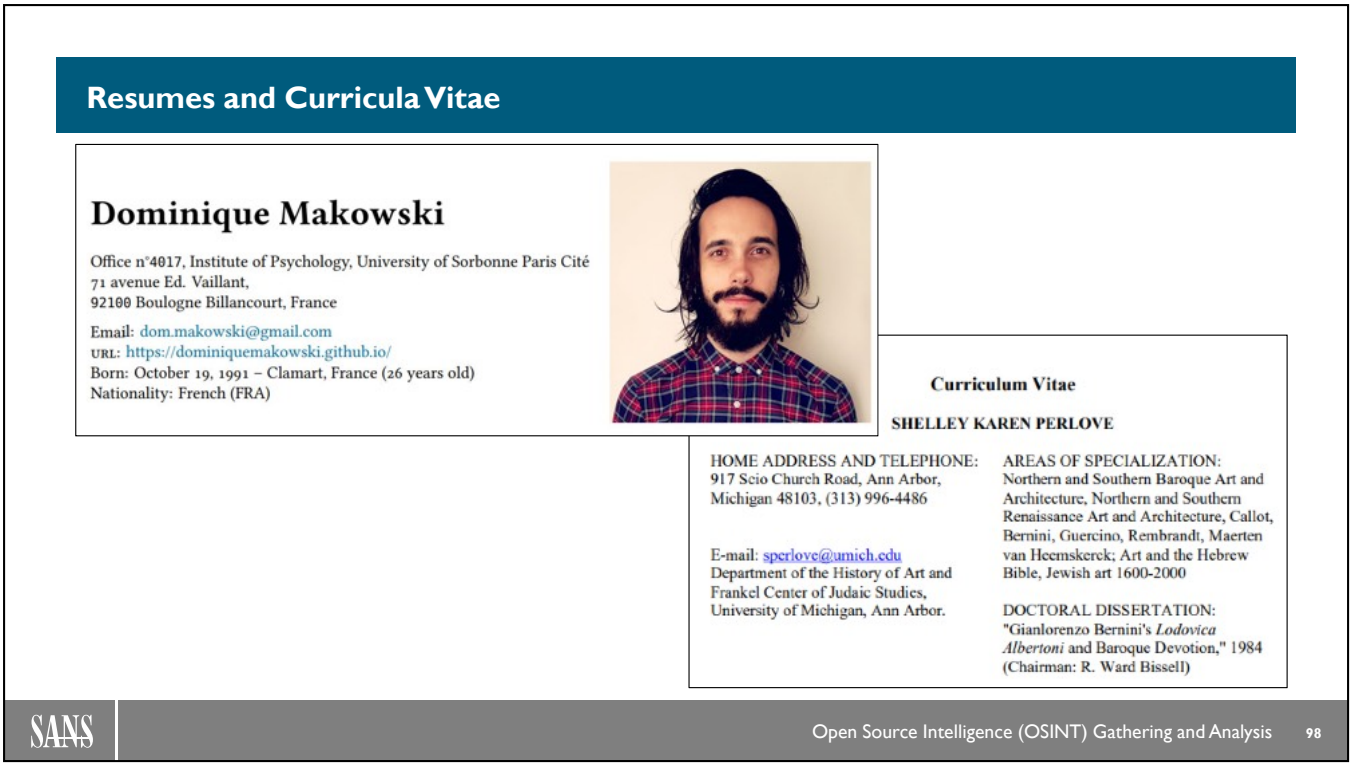
On the results page you can find addresses



People Search Engines

At least within the United States and Canada, there are a variety of "People Search Engines" focused on retrieving data about a person. We cover these search engines in more detail in other places in the course, but understand that many of these sites allow someone to search for a person and the results page will have that person's home address. Aside from the whitepages.com web site shown above, pipl.com, knowem.com, and familytreenow.com (as well as many others) can be used for this type of name-to-address resolution.

Image from https://sec487.info/ed, July 1, 2017.



Resumes and Curricula Vitae

Another method of finding addresses from people's names is to look for documents that contain addresses. A common document that people usually place their name, address, phone number, and sometimes email addresses in is their resume or curriculum vitae (CV). Performing a Google search for `'ext:pdf ext:doc resume OR "curriculum vitae" -sample'` we get over 3,900 results with possible resumes. Obviously, your query may look more like `'Firstname Lastname resume OR "curriculum vitae"'` unless you are looking for resumes from a certain company, in which case you'll replace the first and last names with the company name.

The slide above shows two resumes found on the internet. Both contain name, address, phone, email address, and much more data. As we will see later in the course, resumes can be valuable documents to collect and analyze in your investigations.

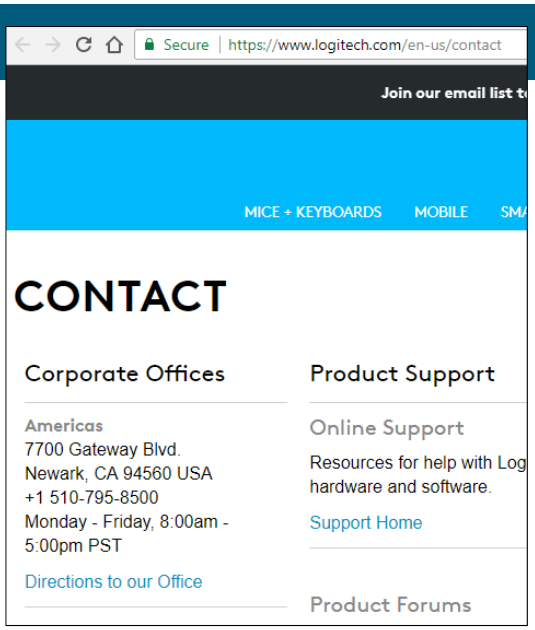
Images from https://sec487.info/kd (July 1, 2017) and https://sec487.info/ke (October 1, 2018).

Company Web Sites

When performing OSINT on a company, you will want to visit their web site (if they have one)

To get addresses, look for pages named:

- About us
- Locations
- Contact us



Company Web Sites

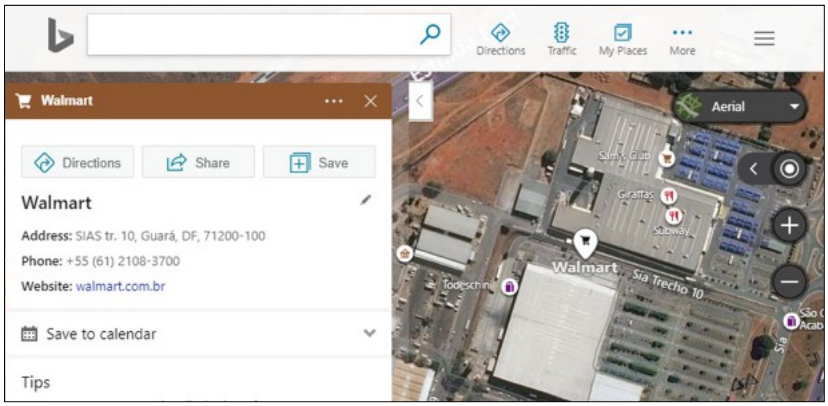
This should not be a surprise, but if you want to get a physical address of a company, visiting that company's web site and looking for pages named "About Us," "Contact Us," and "Locations" is probably going to be helpful to your work. In the example above, we visited the Logitech company (makers of computer accessories) and found a "Contact" page with multiple addresses and phone numbers displayed.

Image from https://sec487.info/eh, July 1, 2017.

Mapping Sites

Many mapping sites allow users to zoom in and click on buildings to find out what their addresses are

Some buildings may be tagged as having certain companies inside them



Mapping Sites

Mapping web sites are designed to give a rich set of information to their users and deliver it graphically. Using your favorite mapping web site (in the example above, we chose to use Microsoft's Bing Maps), you can find a general location. When you wish to get a specific address, either clicking on the map (Google) or right-clicking the map (OpenStreetMap, Bing, and MapQuest) should bring up the data you wish. Many times, you will also see the name of a company or business at that location displayed on the map.

The Bing map above shows the address of a Walmart in Brazil's capital city Brasília.

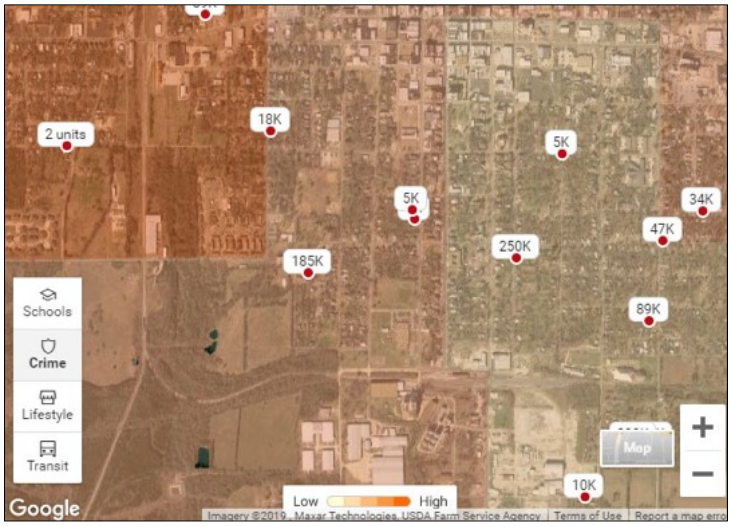
Image from https://sec487.info/kf, December 12, 2018.

Augmenting Data with Real Estate Sites

Real estate sites provide extra information about the addresses you find

Some sites show:

- Interior details
- Crime levels for the area
- Average salaries for homeowners
- Quality of schools



Augmenting Data with Real Estate Sites

Once we find the street address of the location we wish to gather information about, we can gain additional knowledge by visiting real estate web sites. Many real estate web sites contain data about the socioeconomic levels, quality of schools, average temperatures and rainfall, and crime levels in the neighborhood you may be interested in. Within the United States, realtor.com, trulia.com, and zillow.com are three of the many real estate sites that can be used to gather free data once you have an address. Internationally, there are big name sites such as christiesrealestate.com and sothebyshome.com as well as regional sites that are more focused on certain countries.

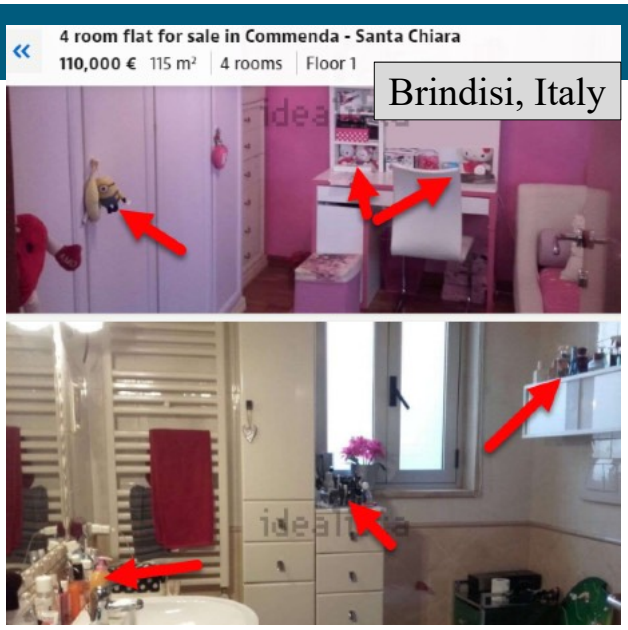
The image above (https://sec487.info/8d) shows the crime tab on the realtor.com web site when you search the Paris, Texas area.

Images from Inside Homes

To "show" a home, pictures are usually taken and published online

Images contain sensitive data about the people who live there

Some are never removed from real estate web sites



Images from Inside Homes

Selecting a random home for sale (posted on the idealista.it web site) in the Brindisi area of Italy, we see interior images taken of what appear to be children's rooms. There are personal effects shown in these images and in others on the web site. The bathroom pictures show what personal hygiene products the owners may use.

These images, since they are posted publicly, may persist on the internet indefinitely after the home has sold and can provide a historic account of previous owners, layouts of the rooms, furniture contents, etc.

Image from https://sec487.info/rg, September 4, 2019.

International Real Estate Sites

International real estate companies share data about homes and neighborhoods too

UK - zoopla.co.uk

International - sothebysrealty.com

Floor plan of home on Sotheby's Site[1]

International Real Estate Sites

Real estate sites around the world also yield useful OSINT data. From images inside homes to floor plans, these web sites can be valuable resources in your investigations. They often provide similar content to real estate sites in the United States, showing neighborhood crimes rates, house values, and accessibility to public transportation.

Reference [1] and image from https://sec487.info/iv, October 10, 2018. (URL no longer active)

Addresses Summary

Addresses can be key pivot points to discover:

- Other addresses (vacation homes, rental properties, etc.)
- People
- Companies

Always perform additional searches on the addresses you collect to discover this information

Addresses Summary

As you have seen in this section, finding a person's or company's address can be an important step towards discovering more information about your targets. Collect and analyze all physical addresses that you find in your work. Seek to obtain addresses from your targets, the places they visit (from geolocations), and the companies they frequent.

Why Gather Phone Numbers

They serve as user accounts

They tie people to property (other offices, vacation homes…)

People and businesses usually do not change phone numbers

Why Gather Phone Numbers

As phone numbers are tied to businesses and people, they provide meaningful links to other bits of data that may be important to our OSINT assessment. Some web sites use phone numbers as user names or as two-factor devices (sending them a PIN via SMS at login). For landlines (non-cell phones), there also may be physical addresses associated with the numbers.

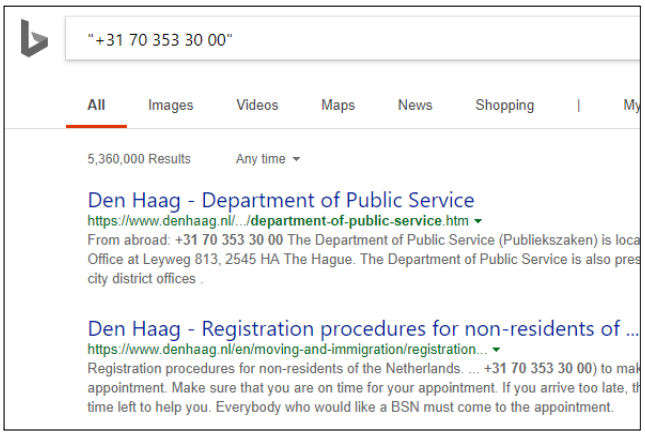
Phone numbers are also fairly unique to people, so much so that many web sites now use them for primary or secondary authentication factors. Some businesses get blocks of phone numbers that they use. With phone number portability (meaning that the number moves with the owner instead of staying with the phone of the carrier), people are keeping their numbers for longer periods of time. Because of this, we sometimes can pull historical data on a phone number and see who uses it now and in the past. This may be of interest or just distracting to your investigation.

Image from https://sec487.info/kg, October 9, 2017.

Using Search Engines with Phone Numbers

Try:

- Grouping the numbers differently
  - `+3170 353 3000`
  - `+31 70 353 30 00`
- Adding quotes
  - `"+31 70 353 30 00"`
- Adding and removing country code



Using Search Engines with Phone Numbers

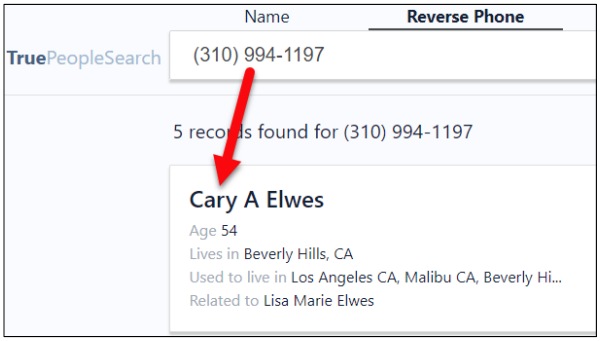
We can use search engines to find information about phone numbers too. When performing these searches, try varying the groupings of the numbers (as shown above) to increase the positive results returned. Additionally, adding quotes around the numbers (especially if there are spaces in them) can help find relevant information. Finally, for international numbers with country codes, perform searches with and without the leading country codes.

Image from https://sec487.info/kh, October 10, 2018.

Reverse Lookups of Phone Numbers

In addition to finding phone numbers, we many times are given them to validate and verify who owns them

Reverse searches (give a phone number, get the owner's data) can open up investigations by providing additional data about the owner



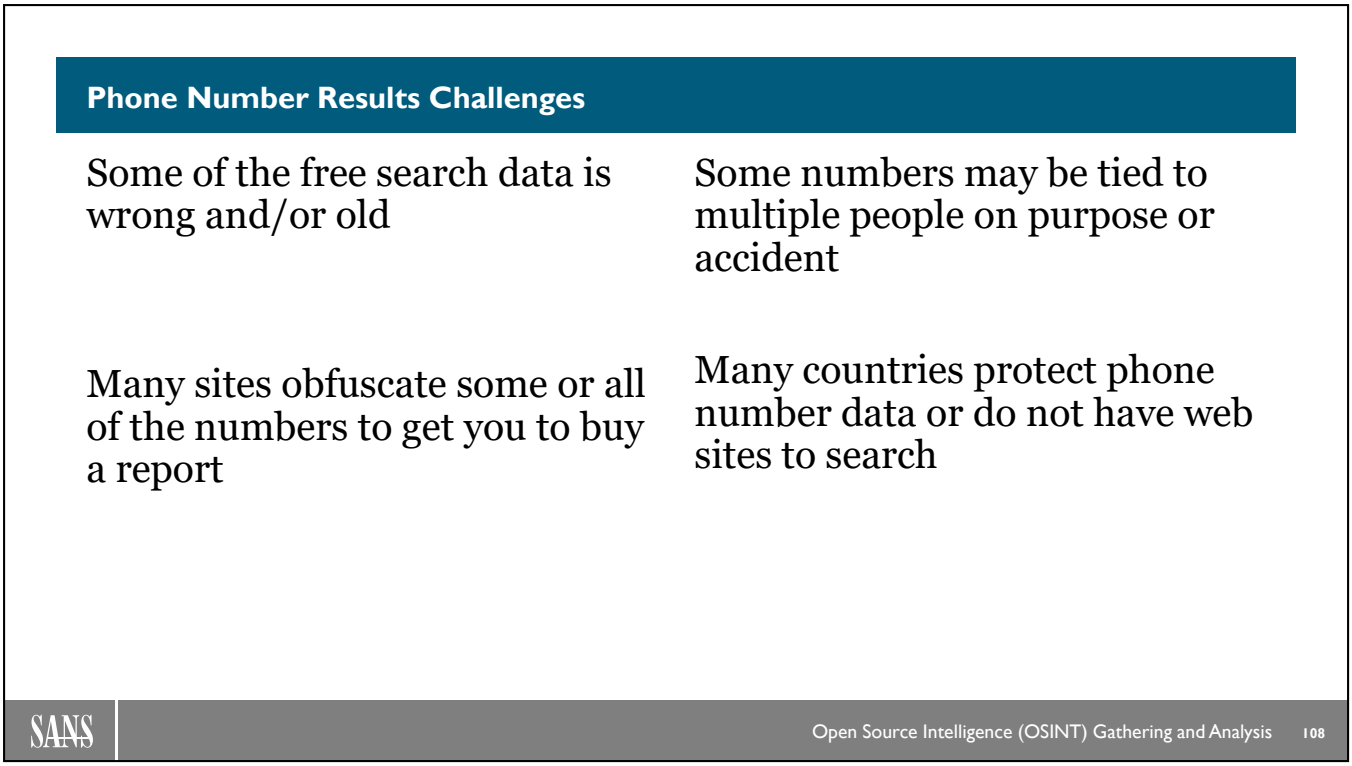
Reverse Lookups of Phone Numbers

Sometimes we have a person's name or an email, and we want to learn their phone number. Perhaps your client is looking for the beneficiary of a life insurance policy, and the phone number they were given in the document was wrong. We may be given a name and a city, state, or country and must find the number(s).

Alternatively, many times our customers provide us with a phone number and ask us to find the owner. Maybe your customer is receiving harassing phone calls from a certain number and they want to find out who it is.

In the slide above, we performed a reverse phone search[1] on the phone number 310-994-1197. The results that came back showed that it may be the phone of *The Princess Bride* actor Cary Elwes (he played Farm Boy Westley and Dread Pirate Roberts in the movie). The next step would be to record and then verify this data and search for other numbers attributed to the target.

Reference and image from https://sec487.info/8x, October 9, 2017.



Phone Number Results Challenges

As with much of the data on the internet, it is only as good as when it was last updated. Depending upon how the information was uploaded or imported into the phone search engines, you may have old and inaccurate data coming out in the results.

Some of the free forward-search person-to-phone-number sites obfuscate either the first or last portions of the phone numbers displayed. Their goal is to tease you into paying for the rest of the digits. OSINT Tip: Use multiple search engines to gather this data, as some blank the first digits and some blank the last. Put them together and you have the whole numbers! Alternatively, use other sites that do not block parts of the data.

There are times when one phone number may be tied to more than one person or business. It could have a business name and the owner's name too. Or perhaps the user of the phone number is not the owner of the bill (think about a child having her parents pay for a cell phone). The account may be in the custodian's name and not the user of the device.

While the United States and Canada have many people, address, and phone number lookup sites on the internet, other countries may or may not have them. This can be limiting, but the privacy policies and laws in some countries may prevent online hosting of this data.

Faking Caller ID Numbers

Trivial to spoof phone numbers using SpoofCard[1] and other systems

This may be illegal in some countries

Work-arounds to use SpoofCard-like apps:

1. Set up USA-based App Store account
2. Use VPN to visit App Store from USA VPN endpoint
3. Download and install "spoof" apps on mobile device

Faking Caller ID Numbers
Most of us understand that the information we might see displayed on a caller ID window on our phones can be faked. There are services such as SpoofCard[1] that make this trivial. In these cases, this could make the phone number that your customer provided to you as seed data inaccurate.

If you need to fake a caller ID number, SpoofCard and those types of services can make it easier to accomplish. If you or your targets are in a country that prohibits falsifying caller ID data, there are methods that you/they could use to do it.

Image from https://www.spoofcard.com/, December 24, 2017.

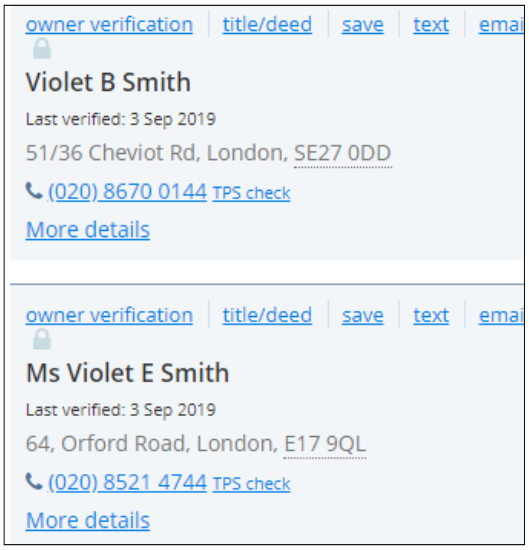
Reference:
[1] https://www.spoofcard.com/

Phone Searching in the UK

Using the ukphonebook.com site, we can search for name, address, and phone number for targets in the United Kingdom

Here we searched for "Violet Smyth" in London

Additional details on the target may be available for purchase



Phone Searching in the UK

Within the United Kingdom, the ukphonebook.com web site, owned by the Simunix company, provides access to some "people" information, including name, phone number, and address, for free. If you buy credits on their site, you can unlock additional details about your targets. With over 130 million records for people in the UK, this site may have exactly what you are looking for. There are other searches that this site offers for additional fees.

Reference and image from https://sec487.info/q-, September 3, 2019.

Phone Number Summary

Phone numbers can be key pivot points to discover:

- Other phone numbers (landlines, mobile, other owners)
- People
- Companies

Always perform additional searches (forward and reverse) on the numbers you collect to discover this information

Phone Number Summary

Phone numbers are a key component to unlocking other details about a person or company. They can be mapped to other people, addresses, or companies. When you find them, perform additional searches, both forward and reverse, to see if you can discover other related data that may be relevant to your OSINT case.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"About My Home"

This page intentionally left blank.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. **Basic Data: Email Addresses**
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.

Why Gather Email Addresses

Format: `local-part@domain`

Example: buttercup@osint.ninja

Unique identifiers for users and tie a person to an activity

The local-part portion may be used as a profile name on web sites

Email addresses **are** user names on many sites, so you can discover other systems the target may use

Email addresses can be used for cyber attacks such as pretexting and phishing

Why Gather Email Addresses

Email has become a ubiquitous method of communicating, so to reach out to someone, you need their email address. If you wanted to email someone named Princess Buttercup, you would need to know what email carrier or domain (examples: outlook.com, gmail.com, protonmail.com, etc.) she uses as well as the local-part[1] of her email, commonly called the email name.

Emails need to be unique so that mail does not get delivered to wrong addresses. Because of this, and because some web sites use email addresses as user names/logins, when we find a target's email address, we have something we can then pivot on and possibly find other accounts and systems they might use.

Finally, once we know a target's email address, if our rules of engagement allow it, we can execute cyber attacks, such as password guessing, phishing, and pretexting, using the email address to achieve our customers' goals.

Reference:

[1] https://sec487.info/91

People Search Engines

Search sites for people inside the United States and Canada are common

Enter the name of your target into the site's search and then narrow down the results using filters

On the results page, you can find email addresses

People Search Engines

When you submit a target's name to them, people search engine sites such as cubib.com and truepeoplesearch.com display email addresses in their results, along with physical address, phone, and other demographic data. These search engines are commonly used with United States and Canadian persons of interest. Looking for someone in another country? Due to privacy policies and laws, you might have a challenging time finding one of these engines that has the information you need.

An interesting OSINT tip is that the user name and local-part of email can be different. Often one informs on the other. For instance, when running the query on the "generic" name of Jane Doe on the cubib.com web site, many of the email addresses had distinct names (for example: sallyjohnson1@gmail.com or Patrick_lewis@aol.com) for the local-part of the email. This may reveal the true user's name.

Images from https://sec487.info/ek and https://sec487.info/ki, July 1, 2017.

Business Email Formats

For consumer email accounts, the email name before the "@" sign could be anything

For businesses, there is usually a pattern to the emails

Most businesses use a combination of user first and last names to form the email

Examples might be:

- first_last@company
- last_first@company
- lastfirst@company
- Flast@company
- lastF@company
- firstL@company
- and so, on …

Business Email Formats

If you cannot find the email of the targets or target organization, then you may need to guess them. If you are looking for a specific target or person, their email address could be anything, and this technique will not work. Most times, people pick some version of their name and something important to them, perhaps a memorable year. An example might be for a user named Kim Lee who got married in 1998 who may have an Outlook account of leekim98@outlook.com.

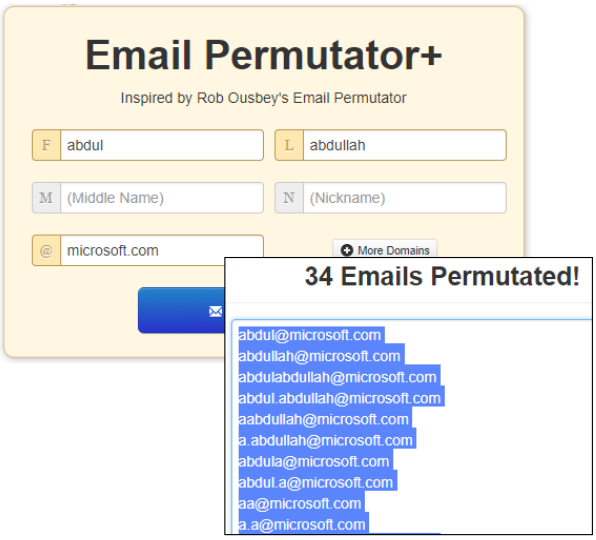
For businesses, there is usually a pattern to most of the user email addresses. This format can be easy to spot on web sites, in forums, and in other places where we may typically find email addresses. The slide above illustrates a variety of combinations of a user's first and last names that companies might use to construct their email addresses.

If we can figure out the format, then we may be able to guess or brute force many of the business email addresses. Depending upon our OSINT goal, this could be extremely useful.

Help with Email Permutations (1)

We could create a script to perform all the permutations of first and last names and the target domain

Or we can use Metrics Sparrow Permutator https://sec487.info/8z



Help with Email Permutations (1)

Using a Python, Ruby, or PowerShell script, we could create all possible email permutations of a corporate email address format using all conceivable combinations of first name, last name, underscore (_), period (.), and the target domain. There are many scripts that will help with this, as we saw about 1.23 million search results from a Google query of "python email permutation" (https://sec487.info/90). Alternatively, we can use a web site such as the metricsparrow.com, as noted in the slide. Simply enter a sample first name, last name, and the target domains into the web form, and the site generates many of the permutations for you.

Images from https://sec487.info/8z, October 10, 2017.

Help with Email Permutations (2)

Or we can use the work from a couple of web sites that have collected many email formats of organizations:

- https://hunter.io/
- https://www.email-format.com/

Help with Email Permutations (2)

We can try to brute force and check all possible combinations of email addresses for a company, or we can use a couple of web sites that may have our target business domain's email format already recorded.

The https://hunter.io and https://www.email-format.com sites are excellent choices to search for your target company. Both contain a large variety of email formats for given companies or domains.

In the slide above from the Email Format site, we see the email formats for the BP company. There are multiple options you can use including viewing identified formats of emails the site has seen, examining the overall format the emails collected have (arrow 1 above), email addresses (arrows 2 and 4), and exporting this data to Excel if you have a paid plan (arrow 3). The hunter.io site not only shows real email addresses, but it also will show you where it found them so that you can examine the original sources.

image from https://sec487.info/qa, August 21, 2019.

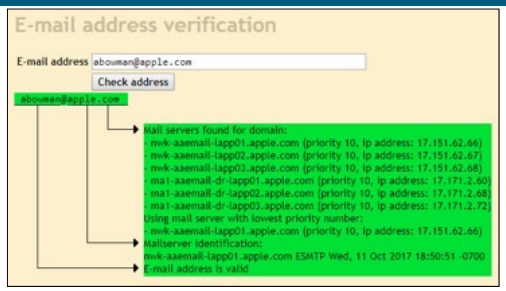
Email Validation: mailtester.com

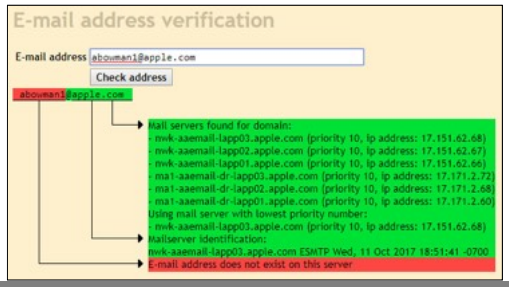
Once we find an email address, we may wish to validate or verify it

This may not matter if it is a user name for some web sites

But if the target might use it for communication, then validate

WARNING: This could tip target





Email Validation: mailtester.com

Finding an email address is pretty simple, especially if you are searching for those from a company that is highly public on the internet. Locating user emails can also be simple if they use systems that display email addresses or use them as user accounts. Once we find them, we might need to verify that they are legitimate. Anyone can make up an email address, such as thisismostlikelynotarealemail.noreally@example.com, and use it as a user name on sites that do not verify email addresses.

Sometimes what we want to do is see if the email is a valid one. For that, we can use a service such as Mailtester.com at https://sec487.info/em. As shown above, this site shows the detailed responses from the remote email server as it checks if an email address you entered is valid. In the example, we found the abowman@apple.com email address on the web and want to verify it. Enter it into the field and submit. Valid email addresses return an all-green response, whereas non-valid email addresses and errors show other colors and text in the response.

OPSEC WARNING: Using techniques to verify email addresses can tip off your target or the company they work for that someone is performing reconnaissance on them. Be extra careful when doing these active measures and make sure it is what your customer wants.

Images from https://sec487.info/em, October 11, 2017.

Gathering Corporate Emails in Bulk

Sometimes your goal will not be to find a single email address but to enumerate as many addresses for a domain as possible

This data could be used for:

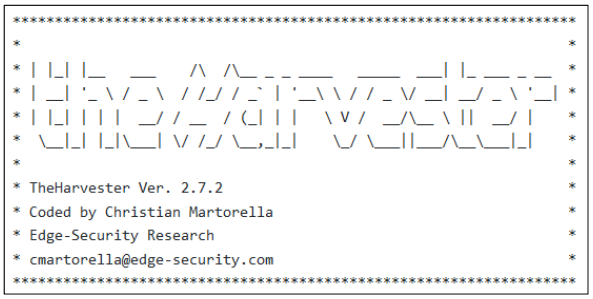
- Phishing targeting
- Social engineering attacks
- Risk assessment (how many of our email addresses are discoverable)

Gathering Corporate Emails in Bulk

Searching for an email address or two on the internet is a common occurrence. There may be assessments where your task is to find many emails from an organization. Perhaps you are performing OSINT to support a penetration test or social engineering campaign. Or maybe you have been asked to find all the people with emails from a certain domain to look for fraudulent entries. We are going to need some tools for this work.

Tools to Get Bulk Emails

There are a variety of OSINT tools that can help harvest a large number of emails (valid or not) about a domain



Examples:

- SpiderFoot
- Recon-ng
- theHarvester

Tools to Get Bulk Emails

When we need to retrieve large numbers of email addresses from internet-based systems, using OSINT frameworks such as SpiderFoot, Intrigue.io, and Recon-ng are critical to working efficiently. Standalone tools such as theHarvester (https://sec487.info/92) can be valuable as well.

These scripts know how to visit a variety of web sites, query them for the information you want (for example, all known email accounts with @apple.com in them), and return the data in a useful format. We cover these tools in depth in other places in the course.

theHarvester

Free Python application written by Christian Martorella

Input: Domain name (-d sans.org)

Sources: Google, Bing, ThreatCrowd, Twitter, Dogpile ...

Output: IP addresses, domain names, email addresses

```
Harvesting results


[+] Emails found:
------------------
info@sans.org
apaller@sans.org

[+] Hosts found in search engines:
----------------------------------

Total hosts: 9

[-] Resolving hostnames IPs...

Computer-Forensics.sans.org:204.51.94.217
access.sans.org:54.71.203.145
blogs.sans.org:204.51.94.130
computer-forensics.sans.org:204.51.94.217
isc.sans.org:204.51.94.153
pen-testing.sans.org:45.60.31.34
retrieve-top-ip-threats-from-httpisc.sans.org:empty
survey.sans.org:23.13.173.28
www.sans.org:45.60.31.34
```


theHarvester

Christian Martorella coded theHarvester, a Python-based OSINT tool (https://sec487.info/92). Provide it a domain name of interest, and it will scour various internet resources looking for IP addresses, domain names, and email addresses that have that domain in them. Check out the web site for the complete list of resources it accesses.

Email Address Summary

Email addresses can be key pivot points to discover

Always perform additional searches on the addresses you collect to discover this information

Emails may have user names in the local-part; pivot on those

Be aware of active email checks and whether they are in scope and good OPSEC

Bulk email searching is easiest to perform with tools

Email Address Summary

Email addresses can be keys to other accounts and are import data points on which to pivot and perform additional searches. Remember that the local-part of emails may divulge something about the user (for personal accounts) or may be used by that user as a valid user name on other web sites.

Use tools and frameworks to facilitate bulk searching and downloading of email addresses, and be wary that, when you verify email addresses for validity, you may be moving from a covert investigation into an overt one.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Finding Emails"

This page intentionally left blank.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. **User Names**
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.

Why User Names Matter

- Many web sites where people submit information require user names
- Tracking target user names can tie them to activities and relationships
- When a user name is created on a website, there is usually a profile with additional information to gather
- The user name itself may have meaning
  - Examples: GoRealMadrid or DreadPirateRoberts1

Why User Names Matter

Usernames are valuable bits of information to collect about our targets. They are many times unique and can allow us to track a single person across multiple sites. Think about yourself. Do you have a single, main user name that you use across your email, gaming, social media, dating, and other web sites? Most people have at least one user name they prefer. If we can discover it, we can use it to find other web sites that our target may have accounts on.

Additionally, when someone creates a user name on a web site, they usually create a profile along with it. Profiles are fabulous places to gather more information about a target.

Finally, we examine the user name itself to see if there is any meaning to the words or characters used.

User Name Uniqueness

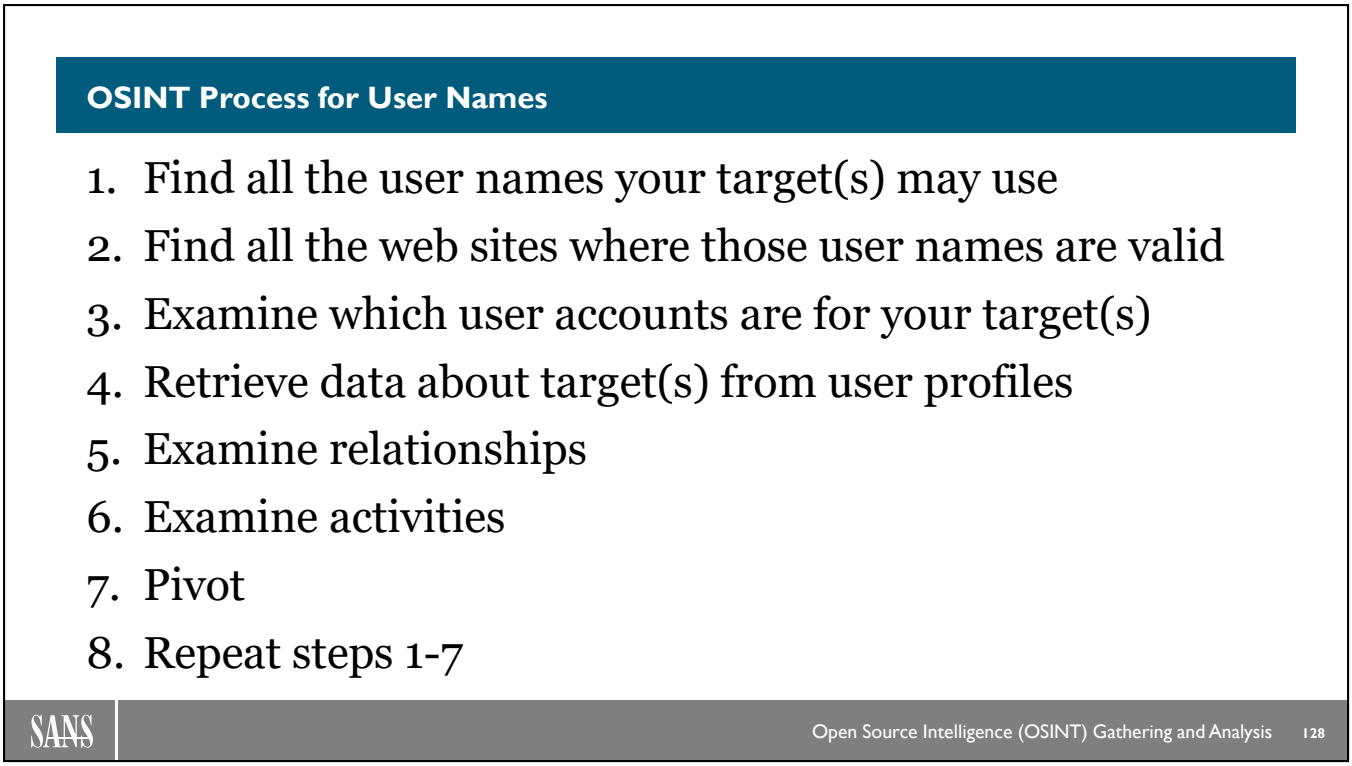
- Most sites require their users to have names unique from other users
- But across several sites…
  - A single user may have the same or different user names
  - Depends on if other people reserved the name first
  - May have mutations of a name
    - DreadPirateRoberts vs DreadPirateRoberts1 vs Dr34dP1r4t3R0b3rts
    - All could be the same person across different sites

User Name Uniqueness

If we understand that people like using the same pseudonym across multiple sites, we can rapidly locate other web sites where our target might have accounts. However, we also need to think about other user name cases. For instance:

- Your target may use multiple user names across a variety of web sites. You must find them all.
- Your target may not have been able to get their favorite user name because someone else reserved it.
- In the above case, you may find that the target has varied their favorite name by adding characters to the end, using a slightly different spelling, or in other ways.

Research based on user name alone and no other corroborating data may have confidence issues because we know that, on most web sites, anyone can have any unique user name. Perhaps you use the user name "DreadPirateRoberts" for your handle in games and also have the email address dreadpirateroberts@example.com. I may love the movie *The Princess Bride,* too, and may have already used that user name for my profile on other sites. Just because we see a unique user name that our target may like does not necessarily mean it is our target's account. We need to look for other corroborating data.



OSINT Process for User Names
When our investigation requires us to search for and analyze user names, we can use the following process to extract as much meaningful information as possible from the data:

1. Find all the user names your target(s) may use – Remember that many people have multiple user names that they use. You need to discover and gather all the names.
2. Find all the web sites where those user names are valid – Once you understand all the names your target(s) may use on web sites, finding all the places on the internet where they have accounts is important so that you can gather profile information, discover activities of interest, and map networks of associates.
3. Examine which user accounts are for your target(s) – Many of the web site profiles you discover could be false positives as, on the internet, anyone can create and use any user name. Our role here is to corroborate the data we found with other information to ensure that each profile hit is actually the target we want.
4. Retrieve data about target(s) from user profiles – Here we examine the "about this profile" pages where many web sites publish biographical data about specific users. Everything could be important, so gather and record all the information on these pages.
5. Examine relationships – Some web applications allow users to connect to other users. Sites may call these connections, followers, friends, or some other relationship type of word. Examine and record all relevant relationship information for later analysis.
6. Examine activities – The user accounts on the web sites most likely have some activities that are published. It may be runs or bike rides they completed, forum conversations they participated in, location "check-ins," or something else. Your job here is to examine, record, and analyze this data for information relevant to your investigation.
7. Pivot – With all the data you collected, choose something relevant to your investigation and then look for activity on that. This may be an important person who your target has interacted with, a location where a certain activity occurred, or something else.
8. Repeat steps 1–7 – With this new item to research, repeat steps 1–7.

Since we as OSINT analysts do not know where the important information will be found and sometimes do not know what is significant until after our analysis, we need to collect everything. The pivoting and repeating are especially important parts of this process. The author has conducted many investigations where the target is connected to another person on a social web site, and that second person provided the necessary information to complete the investigation.

Web Sites to Assist in User Name Checking

There are several web sites that can help us find user names across multiple sites

We will examine two:

- namechk.com
- checkuser.org

All results require validation and confirmation

Web Sites to Assist in User Name Checking

There are several web sites that will take a user name (provided by the user) and query multiple web applications to see if that user account is a valid user on those sites. These services are usually offered to help users select user names that are do not already exist on web sites. As OSINT analysts, we use it in the reverse manner: find out where a specific user name is used and then visit those web sites to gather OSINT data.

Two well-known sites in this space are namechk.com and checkuser.org.

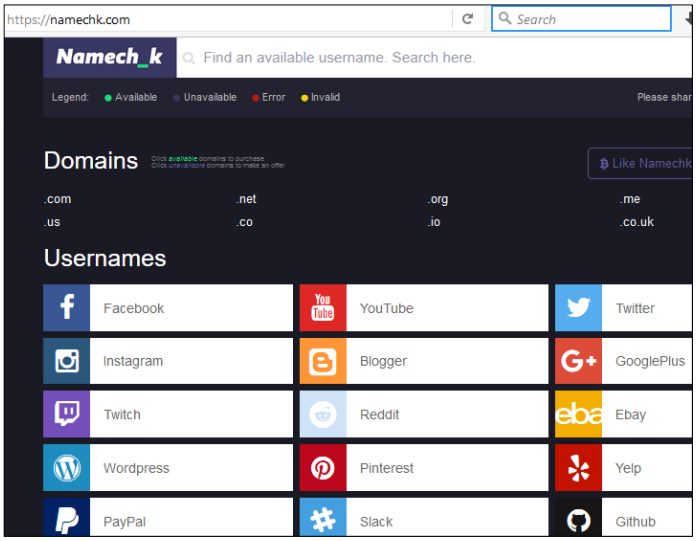
Keep in mind that we do not blindly trust the data on these web sites. The author has encountered many cases where the information returned by one of these sites was false (they either stated that a user name was found on a site and it was not or the reverse). As the OSINT analyst, you must verify and validate the results obtained from these sites.

Namechk.com Intro

Hacker-ish look with black background

Works similarly to knowem.com with server making the requests but also has HTTPS support

Enter user name in search field and go!



Namechk.com Intro

Effective OSINT analysts use multiple sources to both discover data about their targets and corroborate existing information. We can add to what knowem.com found by performing a similar search on the Namechk.com site.

This site has some differences from knowem.com that OSINT analysts should understand:

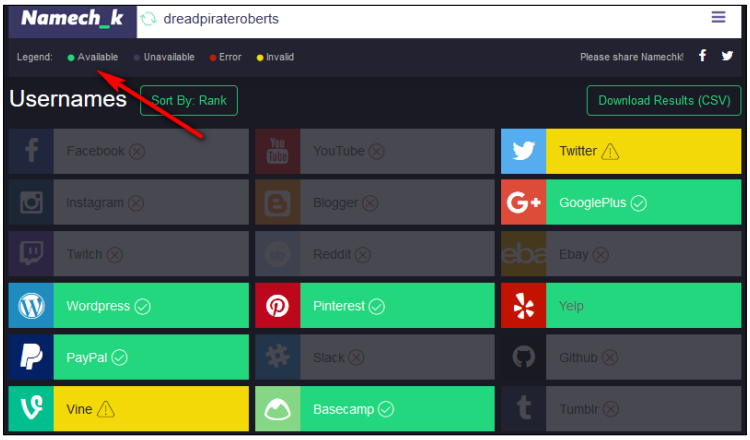
1. The site has a black background so that analysts can feel "hackerish" when performing searches. This also makes the Namechk.com site a "go-to" site to use when your customer is watching everything you do over your shoulder. The black background gives your investigation techniques a "cool factor." ☺
2. This site supports HTTPS communication, which decreases the ability for an adversary or network device to intercept, monitor, and record your queries.
3. While some of the remote web sites that Namechk.com queries are the same as knowem.com, there are some different ones that it searches as well. This gives the OSINT investigator a broader coverage for their investigation.

Image from https://namechk.com/, August 4, 2016.

Namechk.com Results

Results are easy to understand by the background color/shading of each entry

- Green = Available
- Yellow = Invalid
- Red = Error
- Grayed out = Unavailable ← WIN!



Namechk.com Results

The results page that is returned is easy to decipher, as it has a key to the colors of each entry at the top of the page. Green backgrounds for a site show that the user name is available. For our purposes, we are looking for sites where the user name is NOT available because that means that someone, possibly our target, has already reserved the name. So, we discard green-colored results on this page.

Yellow-colored entries are ones where Namechk.com could not search the destination site because the search name you entered was not valid. For instance, Twitter handles (names) can be up to 15 characters long (https://sec487.info/w). So, searching for "dreadpirateroberts" would not be a valid search because no user could have that Twitter handle.

If Namechk.com encounters an error when searching a system, it will color an entry with a red background. We can either ignore these entries or manually verify them.

You will need to record all the grayed-out entries on the results page, as those are the places where the user name you searched for was already taken. Just as with knowem.com, this will be a long and boring process. Is there an easier way?

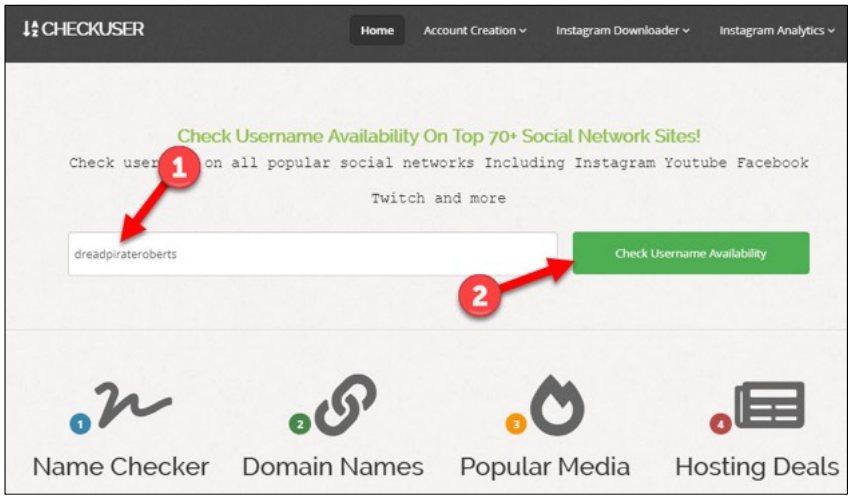
Image from https://namechk.com/, August 4, 2016.

checkuser.org Site

Simple interface

Enter username (1) and click button (2) to search

This web site makes requests to the sites to request user names



checkuser.org Site

Another site worth mentioning in this space is checkuser.org. This web site has a simple search field where you enter the user name you wish to search for. Click the search button, and your search has begun. It is worth noting that the web site does all of the requests to the social media and other sites that are being checked instead of those requests coming from your web browser and IP address. This buffers your traffic, so the target web sites don't know that you, from your IP address, are interested in a certain name.

Image from https://checkuser.org, August 20, 2019.

checkuser.org Results

Grayed-out results (arrow 1) are what we want

Black font results need to be checked or cannot be due to length (arrow 2)

Green font can be researched last (3)

Arrow 4 points to scroll bar to see more results

checkuser.org Results

Similar to namechk.com, checkuser.org shows results the way we would expect them: sites that HAVE accounts on them with the target user name are grayed out (arrow 1). Sites with no account with the name we requested have a green font and say "Available."

This web site also presents us with the URLs to where we can go and look at what it found when it checked each site. This helpful step makes this site something that is easier to use for our purposes.

One drawback to using this and the namechk.com sites is that we still don't know how they figured out if there was or was not a user account on each site.

Image from https://sec487.info/q9, August 20, 2019.

Using Google to Find User Names

- Google has the "inurl:" modifier that will search URLs for content
- We can also use the word "OR"
- We can use something like the following in a Google search box to look for URLs that have the word “profile” or “user” in them and also the name dreadpirateroberts
  - `inurl:profile OR user inurl:dreadpirateroberts`
- You may need to alter the names or terms for your query

Using Google to Find User Names

The Google search engine allows users to search the URLs it has indexed. This can be helpful to OSINT analysts, as sometimes web application URLs contain keywords such as "profile" and "user." By crafting our queries using search modifiers (which we will go into more in depth later today) such as "inurl:”, we can find places where Google has indexed web pages with our terms in the URL. Google also allows us to use the modifier "OR" to tell it that we want one term or another. So, the terms "soup OR salad" would tell Google to search for entries with the word "soup" in it or the word "salad."

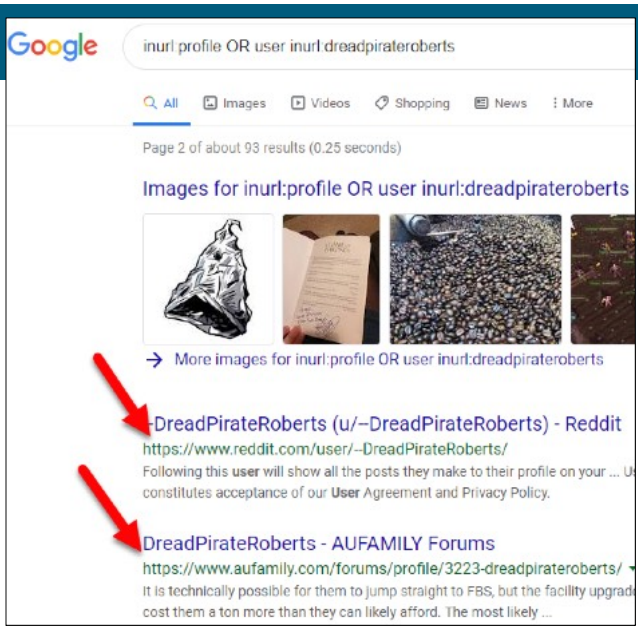
Some web sites use the keywords "profile" and "user" for places in their sites that contain user account information. If my web site had the URL "https://example.com/users/DreadPirateRoberts" and Google searched it, then we can use the "inurl:" modifier to find our target.

We will start with a basic query of "`inurl:profile OR user inurl:[USERNAME]`”, where you replace [USERNAME] with the user name of your target. In our example, we can use "`inurl:profile OR user inurl:dreadpirateroberts`”, which tells Google to search for pages with the word "profile" or "user" in it and also where "dreadpirateroberts" appears in the URL. This query in Google can be reached through https://sec487.info/x.

Google User Name Search Results

Using the search terms on the previous page gives us results that may show web sites that we had not known about

In the picture on right, examine the URLs to find other user names containing DreadPirateRoberts name and numbers and dashes (--)



Google User Name Search Results

Once we perform the Google search from the previous slide, our results may show us known URLs or they may provide us other web applications where our target's profile name was located. They may also give us variations of the user name, which we can leverage to gather more data about our target. In the search results above, we can see that, of the results, some of the user account names contained variations of the name ("--DreadPirateRoberts" and "3223-dreadpirateroberts").

As OSINT analysts, we will record these possible user names and the sites where they were located and then visit them to validate that the account(s) belong to our target.

Image https://sec487.info/x, August 20, 2019.

More Efficient User Name Searching

- The problems with user name enumeration is not the searching of target sites but what comes after
  - Manually recording sites where a target user name appears
  - Trying to figure out how to find the correct URL to access the site's profile area
- This is challenging and time-consuming
- So, the "WhatsMyName" project was created

More Efficient User Name Searching

We have seen several web sites that will quickly search a wide variety of sites to find a user name of interest. The problem that we have is that there is a lot of work to do by the analyst once the results are returned. Remember, these sites take a single user name. If you have a gang or group of targets, this process will be very cumbersome to your analysts.

Another issue we have seen is that, although the web sites have returned whether a given site has a specific user account on it, they do not show the analyst where that user's profile resides on the site. So, for each site, the analyst will need to hunt for where the user's profile content is found. Again, multiply this by the number of sites, and by the number of target usernames, and this quickly becomes overwhelming.

To address these (and other) issues, the WhatsMyName project was created.

WhatsMyName Intro

Micah Hoffman (@WebBreacher), SEC487 course author, discovered many URLs where user name enumeration was possible

The difference with this project is that it has the expected responses for valid and invalid user accounts

Let's see how this project can help the OSINT analyst

WhatsMyName Intro

Recognizing the shortfalls of the knowem.com and Namechk.com web sites, Micah Hoffman (@WebBreacher) created the WhatsMyName open source project. The project is on GitHub at https://sec487.info/l.

Micah took the concepts that knowem.com and Namechk.com use to perform their user enumeration and documented them for anyone to use. The file containing this information is in the JSON (JavaScript Object Notation) format, which is easily decoded and used by humans and computer programs. He and other contributors to the project noted what the web server responses were when valid and invalid users were requested from certain sites. There are over 100 sites in this project.

WhatsMyName JSON File

web_accounts_list.json file contains:

- Site name
- URL
- Valid/invalid codes
- Valid accounts
- Categories*
- Validity


```
},
{
   "name" : "Instagram",
   "check_uri" : "https://www.instagram.com/{account}/",
   "account_existence_code" : "200",
   "account_existence_string" : " Following, ",
   "account_missing_string" : "Page Not Found ",
   "account_missing_code" : "404",
   "known_accounts" : ["test", "barackobama"],
   "category" : "social",
   "valid" : true
},
{
   "name" : "instructables",
   "check_uri" : "http://www.instructables.com/member/{account}/",
   "account_existence_code" : "200",
   "account_existence_string" : "joined",
   "account_missing_string" : "",
   "account_missing_code" : "404",
   "known_accounts" : ["test"],
   "category" : "hobby",
   "valid" : true
},
{
   "name" : "Internet Archive",
   "check_uri" : "http://archive.org/search.php?query={account}",
   "account_existence_code" : "200",
   "account_existence_string" : "item-ia",
   "account_missing_string" : "",
   "account_missing_code" : "200",
   "known_accounts" : ["test","mubix"],
   "category" : "search",
   "valid" : true
},
```


WhatsMyName JSON File

The `web_accounts_list.json` file has all of the important data that we need to find our user names on over 150 different sites. The README.md file contains a good description of the file, explaining that each entry can have the following fields:

```
"name" : "name of the site",
"check_uri" : "URI to check the site with the {account} string replaced by
               a username",
"pretty_uri" : "if the check_uri is for an API, this OPTIONAL element can
                show a human-readable page",
"account_existence_code" : "the HTTP response code for a good 'account is
                            there' response",
"account_existence_string" : "the string in the response that we look for
                              for a good response",
"account_missing_string" : "this OPTIONAL string will only be in the
                            response if there is no account found ",
"account_missing_code" : "the HTTP response code for a bad 'account is not
                          there' response",
```

```
"known_accounts" : ["a list of user accounts that can be used to test",
                    "for user enumeration"],
"category" : "a category for what the site is mainly used for",

"valid" : "this true or false boolean field is used to enable or disable
           this site element"
```

* Some of the URLs that are in this project are for dating and pornographic and sex web sites. The reasoning behind this is that sometimes people share content in these web pages that may be embarrassing and may be important to our investigations. These entries will be marked "XXX PORN XXX" in the results. If you or your company do not wish to see the content of these pages, do not visit them. However, realize that, unless you remove those entries from the JSON file, any script that uses this list WILL make requests to those sites.

** Using this list to make web calls behind a web proxy such as McAfee's Web Gateway, Bluecoat, or Untangle may cause false positives, as those systems may prevent the DNS lookups for some of these sites and may also block web traffic.

Image from https://sec487.info/kj, September 6, 2019.

WhatsMyName's Purpose

- WhatsMyName provides us a method to automate user name enumeration on a fast scale
- Using a Python, Ruby, Perl, or Bash script, we can quickly make calls to web sites and look for responses in the pages to tell if the user name is valid or not
- We also get the exact URL to the user's profile page, which we did not get from knowem.com and Namechk.com
- So now we need to do programming, right?

WhatsMyName's Purpose

As mentioned previously, the WhatsMyName project was made to make it easier for people to enumerate valid user accounts on a variety of web sites. Because the main data is in JSON format, it is easily leveraged by Python, Ruby, Perl, or Bash scripts in an automated fashion.

A benefit of using this content is that the contributors of this project have done the work for you. When a valid account is located for a given resource, the result will be a URL directly to the user's profile page. This is a huge timesaver for the busy OSINT analyst, as you now do not need to search through site after site of pages trying to find where all the content you care about is located. This is also how the project is differentiated from the knowem.com and Namechk.com sites.

Do not be concerned if you do not know a programming language. The open source community has done that work for you! Of course, if you do know a scripting language, you can create your own script using this data.

Others Have Already Written Scripts

There are at least two main projects that leverage the WhatsMyName JSON file:

- Steve Micallef's (@smicallef) SpiderFoot Accounts Module
  - Explained at https://www.spiderfoot.net/
- Micah Hoffman's Recon-ng Profiler Module
  - Explained at https://sec487.info/n
  - Code included in the Recon-ng.com project

Others Have Already Written Scripts

As of this writing, there are two main projects that use the WhatsMyName data: the SpiderFoot project led by Steve Micallef (@smicallef) and the Recon-ng Profiler module written by Micah Hoffman.

The SpiderFoot project (https://www.spiderfoot.net/) is a general OSINT and Threat Intelligence automation system that can retrieve much more than merely user accounts on remote web sites. Directions, features, and more can be found on the main web site. Code for this project is downloadable from GitHub at https://sec487.info/m.

The Recon-ng Profiler code is a module within the Recon-ng reconnaissance framework, with a full tutorial on module use at https://sec487.info/n. The Recon-ng project is also a general OSINT and threat intelligence tool.

Recon-ng Profiler Module

- The Profiler module accepts users name from Recon-ng and uses the WhatsMyName JSON file to perform user enumeration
- Multi-threaded and can be very fast
- Results are stored in Recon-ng's local database
- Requests are made from the system Recon-ng is running on, NOT a dedicated server

Recon-ng Profiler Module

The Profiler module take a user name or names from the Recon-ng database and leverages the WhatsMyName JSON file URLs to make requests. These web calls are made from the system that Recon-ng is installed upon (unless you are using an advanced technique such as tunneling traffic or using a VPN). With Recon-ng's multi-threaded execution and a fast internet connection, it is easy to make hundreds of user name enumeration requests in under one minute. The author ran three user names through the module and had it check 190 web sites for each account, and had the results in 30 seconds. It is that fast!

The user names that are used in the module are pulled from the Recon-ng's local "profiles" database table, and results are stored in the same place. If you have not used other modules in Recon-ng for reconnaissance, discovery, or exploitation, the "profiles" database may be empty. We will need to add manual entries to it before executing our module.

Recon-ng Profiler Setup

How to run the Profiler module interactively

1. Launch Recon-ng from a terminal
2. Select or add a "workspace" for your investigation
3. Use the profiler module
4. Add the user name to the profiles database table
5. Add another user name if desired
6. Run the module

Recon-ng Profiler Setup

After downloading Recon-ng and resolving all dependencies, you will need to run the program. Recon-ng has several methods of doing this. We will show the interactive method here.

Launch Recon-ng interactively and disable the Google Analytics by typing:

```
./recon-ng --no-analytics [press enter]
```

Add a workspace named "test1" by typing:

```
workspaces add test1 [press enter]
```

Use the Profiler module by typing:

```
use profiler [press enter]
```

Add user names to the profiles database table by typing (note the 4 tildes (~) on the end):

```
add profiles dreadpirateroberts~~~~ [press enter]
```

Add other user names to the table by repeating the above command and inserting other names.

Run the module by typing:

```
run [press enter]
```

Recon-ng Profiler Setup (Displayed)

Recon-ng version 5.x

```
[recon-ng][test1][profiler] > workspaces create test1
[recon-ng][test1][profiler] > modules load profiler
[recon-ng][test1][profiler] > db insert profiles dreadpirateroberts~~~~
[recon-ng][test1][profiler] > db insert profiles farmboywesley~~~~
[recon-ng][test1][profiler] > run
```

Recon-ng version 4.x

```
[recon-ng][test1][profiler] > workspaces add test1
[recon-ng][test1][profiler] > use profiler
[recon-ng][test1][profiler] > add profiles dreadpirateroberts~~~~
[recon-ng][test1][profiler] > add profiles farmboywesley~~~~
[recon-ng][test1][profiler] > run
```


Recon-ng Profiler Setup (Displayed)

Here is a picture of what the commands on the previous page should look like when executed. There are alternatives to adding items to the profiles database in this manner. If you have a single user name to examine, you can "set SOURCE username" in the module.

For example, if we only wanted to search for the "dreadpirateroberts" user, once we get into the Profiler module in Recon-ng:

- In Recon-ng version 5, that command changes to `options set SOURCE dreadpirateroberts`
- In Recon-ng version 4 we could use `set SOURCE dreadpirateroberts`

Recon-ng Profiler Module Results Output

rowid	username	resource	url	category
1	dreadpirateroberts	Canva	https://www.canva.com/dreadpirateroberts	business
2	dreadpirateroberts	BuzzFeed	https://www.buzzfeed.com/dreadpirateroberts	social
3	dreadpirateroberts	cHEEZburger	http://profile.cheezburger.com/dreadpirateroberts	hobby
4	dreadpirateroberts	Disqus	https://disqus.com/by/dreadpirateroberts/	discussion
5	dreadpirateroberts	Etsy	https://www.etsy.com/people/dreadpirateroberts	shopping
6	dreadpirateroberts	FurAffinity	https://www.furaffinity.net/user/dreadpirateroberts	XXX PORN XXX
7	dreadpirateroberts	Bandcamp	https://bandcamp.com/dreadpirateroberts	music
8	dreadpirateroberts	GitHub	https://api.github.com/users/dreadpirateroberts	coding
9	dreadpirateroberts	Garmin connect	https://connect.garmin.com/modern/profile/dreadpirateroberts	health
10	dreadpirateroberts	Gravatar	http://en.gravatar.com/profiles/dreadpirateroberts.json	images
11	dreadpirateroberts	Geocaching	https://www.geocaching.com/profile/?u=dreadpirateroberts	social
12	dreadpirateroberts	HackerOne	https://hackerone.com/dreadpirateroberts	hacker
13	dreadpirateroberts	IFTTT	https://ifttt.com/p/dreadpirateroberts	misc
14	dreadpirateroberts	Instagram	https://www.instagram.com/dreadpirateroberts/	social
15	dreadpirateroberts	Flickr	https://www.flickr.com/photos/dreadpirateroberts/	images
16	dreadpirateroberts	Kik	https://kik.me/dreadpirateroberts	social
17	dreadpirateroberts	LibraryThing	http://www.librarything.com/profile/dreadpirateroberts	books
18	dreadpirateroberts	MassRoots	https://api.massroots.com/v1/accounts?singleUser=true&username=dreadpirateroberts	drugs
19	dreadpirateroberts	meet me	https://www.meetme.com/dreadpirateroberts	dating
20	dreadpirateroberts	Medium	https://medium.com/@dreadpirateroberts/latest	news
21	dreadpirateroberts	Mixcloud	https://www.mixcloud.com/dreadpirateroberts/	music
22	dreadpirateroberts	Pinterest	https://www.pinterest.com/dreadpirateroberts/	social
23	dreadpirateroberts	Periscope	https://www.periscope.tv/dreadpirateroberts	video

Recon-ng Profiler Module Results Output

While the script runs, it will show colorized results (if your terminal allows) of what it is finding. Positive results, where a user name has been located, appear with a green [*]. Errors are shown in red. Normal status shown in the default color for your terminal.

Once the module has finished running through all of the sites for all of the user names, it will end. If you have positive results, they will have been added to the Recon-ng database and are accessible by typing `show profiles` and pressing Enter. The results for the profiles table will be displayed in table format with each result showing:

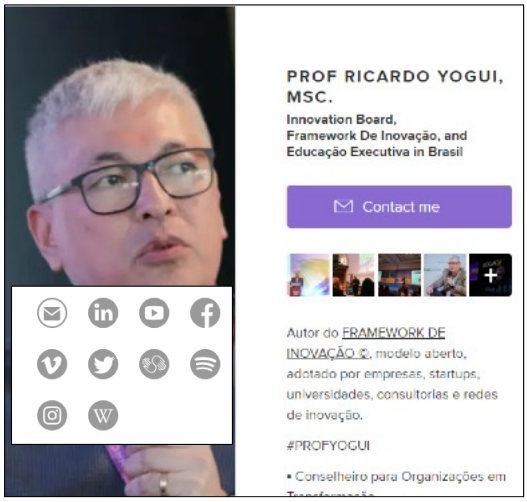
- A unique ID per entry
- The user name that was found on a specific site
- The site name where the positive result was discovered
- The exact URL that can be used to view the profile of the user on that site
- A category of what type of data the site contains

With this data, the OSINT analyst can rapidly discover additional profiles, web sites, and content to explore and enhance their investigation. Remember that this tool shows valid accounts on sites. It is up to the analyst to validate that the user account found is the actual target.

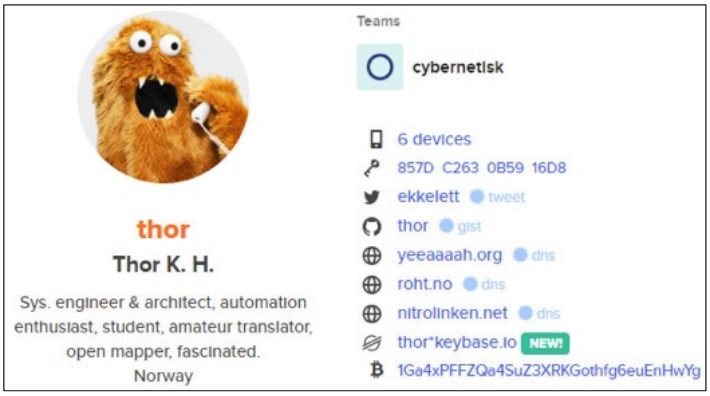
Above is the output from the profiler module after we ran it against the "dreadpirateroberts" account. This run of the script pulled 39 profiles, which would then have to be researched by the analyst.

Tying Social Media Accounts Together

https://about.me



https://keybase.io



Tying Social Media Accounts Together

You, as an OSINT analyst, may not need to work hard to link up a target's social media accounts. They may have done the work for you using a social media aggregator site like about.me or keybase.io. These sites allow users to claim ownership and verify that they are the person who uses certain accounts and web sites. Finding a user's profile on one of these sites can help you rapidly discover related accounts to the target.

The about.me profiles are quite simple to harvest data from. Using your favorite search engine, search for information that may be in your target's profile, along with the "site:about.me" term, and you should find results if they are posted. Remember that people WANT you to find these accounts, so they make them public and easy to find.

Images from https://sec487.info/yh and https://sec487.info/yi, October 5, 2019.

Keybase Foundations

Keybase.io allows users to cryptography connect their accounts and web sites

Only the user knows the passwords for their keys

They verify themselves through posting a string of characters or adding records to DNS

Keybase Foundations

Keybase is a web site (https://keybase.io), a chat client, and a growing set of other tools that reveals public data about its userbase. A person signs up for an account, creates one or more cryptographically secure keys, publishes them on the site, and then uses those to validate their other accounts and websites they control on the internet or to encrypt data to other users. To validate that they control an account, they must tweet or post certain content that Keybase's servers read and validate.

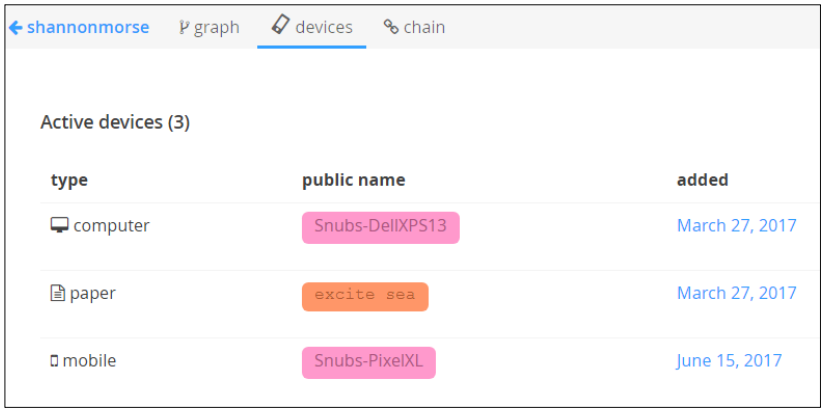
Without downloading any software, the public data this site contains about its users is an OSINT analyst's dream! Users connect their different usernames, web site addresses, and bitcoin addresses for us, saving us much work.

User-Submitted Computer Names

The user "shannonmorse" created device keys for systems named:

- Snubs-DellXPS13
- Snubs-PixelXL

What would the computer names look like from a company?



User-Submitted Computer Names

When a Keybase user begins creating and using their keys, they do so from devices like phones, tablets, and laptops. Each of these devices has a public name that can reveal data about the device itself. As an example, examine the https://sec487.info/hm device page from the user Shannon Morse. If the data can be trusted, Ms. Morse may own or use a Dell XPS13 laptop and a Google Pixel XL phone. We also see a pseudonym, Snubs, as part of those device names. Finally, we see when those devices were added to Keybase. If your customer wanted to know when Ms. Morse got her Google Pixel XL phone, this data could be useful.

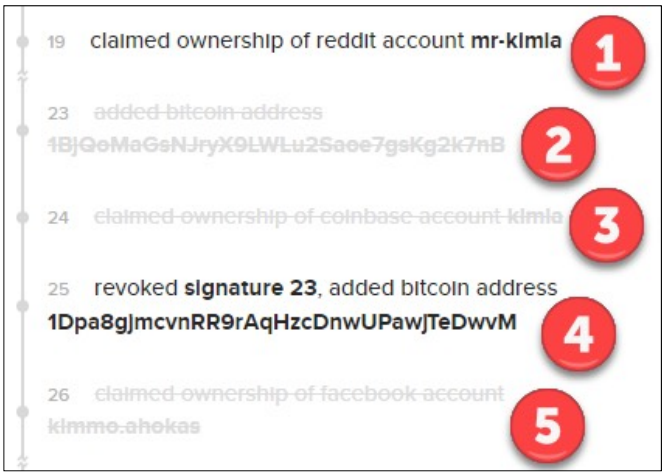
Imagine if a Keybase user used the site from a work computer and allowed the internal computer name of their system to be put into the Keybase data. We may get naming patterns, locations, and employee IDs from that content.

Keybase Signature Chain

Each transaction on Keybase is recorded in a signature chain

This is public and shows:

- Domains
- Computer/phone names
- Accounts the user has control over
- Dates actions were taken



Keybase Signature Chain

As Keybase's documentation notes, "*Every account on Keybase has a public history. 'Sigchains' let Keybase clients reconstruct the present without trusting Keybase's servers. And when you 'follow' someone on Keybase, you sign a snapshot of your view of the claims in their sigchain."* [1] The process to validate transactions and history while not trusting Keybase's servers is a noble and important one for this system. For OSINT purposes, it can be a goldmine of data, as you can see above with the user "kimia," [2] who shares two bitcoin wallets (2 and 4), a Coinbase account name (3), a Reddit account (1), and a Facebook account (5) with us in this sigchain page.

When a Keybase user makes a mistake, perhaps allowing their company's computer name to be used for the public name of their device, they can delete that action so that a person checking out their current list of devices no longer sees that internal computer name. However, every key-based action is stored in the signature chain. As shown in the above slide, we can find data about when the user added devices, verified accounts and sites, and revoked keys.

Image from https://sec487.info/yj, October 5, 2019.

References:

[1] https://sec487.info/mh, March 31, 2018

[2] https://sec487.info/yj, October 5, 2019

Keybase Public API

The unauthenticated, public Keybase API reveals data about a specific user

Replace USERNAME with your target's at the end of this URL:

`https://keybase.io/_/api/1.0/user/lookup.json?username=USERNAME`


```
"them": {
  "id": "75912fb26f7d2b7e2a5bab4f1602da19",
  "basics": {
    "username": "dutch_osintguy",
    "ctime": 1520155722,
    "mtime": 1565618750,
    "id_version": 22,
    "track_version": 23,
    "last_id_change": 1564731793,
    "username_cased": "Dutch_OsintGuy",
    "status": 0,
    "salt": "7c9f8c790808b0294b2db023c98fcd23",
    "eldest_seqno": 1
  },
  "profile": {
    "mtime": 1540111993,
    "full_name": "Dutch_OsintGuy",
    "location": "Netherlands",
```


Keybase Public API

The Keybase site has a free, public API that can be used to query information about targets. In the above example, we get JSON-formatted content back when we add the name "dutch_osintguy" to the end of the query URL: https://keybase.io/_/api/1.0/user/lookup.json?username=dutch_osintguy (https://sec487.info/mi). This API pulls profile information, user devices configured to use Keybase, public keys, and whatever "proofs" the user submitted to prove ownership of various accounts and domains.

Image from https://sec487.info/mi, October 5, 2019.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Finding Users"

This page intentionally left blank.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. **Avatars and Reverse Image Searches**
9. Additional Public Data
10. Creating Sock Puppets

This page intentionally left blank.

Avatars

- Personalized images used on web sites
- Found in social media, forums, comments, reviews, and other places
- Can be unique or common picture
  - Art (beach scene, sport team logo)
  - Photos (family and friend's images)
  - Computer-generated image (random)
- Retrieve and store all avatars

Avatars

Social media sites, forums, and other places on the internet allow users to use an image or avatar to represent them in their online activities. Avatars make it easier for people reading websites to identify that certain content was posted by one user versus another by appearing next to the content that they post. When we tweet on Twitter, our content also shows our avatar. When we post on a social media, dating, or job site, our avatars are shown along with the data we post.

On some web sites, default avatar icons are given to users and may be unique. Users have the option to customize their avatars by uploading or choosing other pictures.

Retrieve and store all avatars according to site for processing and analysis.

Avatars for OSINT

We examine avatar pictures for what they show

- Photos of your target? Their families?
- What other things are in the picture?
- Where was the photo taken/taken from?

Are the avatars used, reused across multiple sites by the same person?

Are other people also using that avatar?

- Possible false positive when looking for a target
- Showing common beliefs, thoughts

Who are the friends/connections of this avatar?

Avatars for OSINT

When conducting an OSINT investigation, the avatars used by our targets can provide important information. They can lead to new avenues to explore, hint at hobbies and interests, and can tie that person to accounts and activities at other web sites. Many times, user-chosen avatars have meaning to the user. Perhaps it is a photo of a person they like or a "selfie". Maybe the avatar represents an organization they are a "fan" of such as a sports team or company logo. As analysts, our roles are to retrieve the images and analyze them to see if there is meaning to the pictures. If there is, this may represent a pivot point (another place to search) in your investigation.

Effective analysts examine the images targets use and ask questions that will further their investigations. Some questions you might wish to look into are:

- What do the avatars show? Are they photos of friends and relatives of the target?
- What other items, scenery, and people are in the picture? Sometimes it is the "other" things in the image that can be important.
- Where in the world was the photo taken? Geographic locations can be important in investigations.
- Was it taken from another web site? Go to that site and examine it for links to the target and content.
- Is the target reusing avatars across multiple sites? We will examine this later in the course.
- Is the avatar a common image that people other than your target are using? This might give you false positives in your searching.
- Does this avatar show up on other people's social media pages? This may indicate a relationship that needs further examining.

Relationships Through Avatars

When we view a social media site, we see can the activities of our target and sometimes of people they are connected to

- Facebook = "Friends"
- LinkedIn = "Connections"
- Twitter = "Followers"

If our target avatar appears in another person's activity feed, we need to analyze it to identify any possible relationships

Relationships through Avatars

Search engine results may also indicate that the target avatar appeared on a friend, connection, or follower's page. This is a common occurrence and should be documented in the analyst's notes. These relationships can be important to investigations and provide pivot points to family, friends, associates, and coworkers. While not necessarily the focus of the investigation, the information obtained from associations should be analyzed. Many times, the target's friends and family will be less security minded, post images of the target, and aid your investigation.

Identify and document any possible relationships, as those become additional avenues of investigation.

Avatars Across Sites

Gravatar (https://en.gravatar.com/)

- "A **G**lobally **R**ecognized **Avatar**"[1]
- Upload once, and other sites pull from Gravatar.com
- Same avatar for a given user across multiple sites

Quick URLs to access human-readable profiles

- http://[language].gravatar.com/profiles/[user name]
- Example: http://ru.gravatar.com/profiles/dreadpirateroberts1
  - Shows dreadpirateroberts1's profile in Russian (try en, fr, de, and others!)
- Other formats available such as XML and JSON
  - Append a .json or .xml extension to the end of the above URLs

Avatars Across Sites

The Gravatar web site provides a wonderful service to the internet community. Instead of having to upload your same picture and profile data to every site where you create an account, if the sites you are using supports it, you can create an account and upload your data once to Gravatar, and each site can pull that content.

Gravatar profiles are created using user accounts at Wordpress.com.

Accessing a target profile is easy using a web browser or command-line tool. The Gravatar site is available in many languages and can be extremely helpful in tying a target's data together and augmenting information you have already acquired. To get the site to appear in a specific language, use a two-letter code (for example, en=English, fr=French…) and substitute it in the server name portion of the URL. An example is viewing the profile page for the "dreadpirateroberts1" account in Russian using the http://ru.gravatar.com/profiles/dreadpirateroberts1 URL. Note the "ru" in the URL stands for Russian. For other languages, see https://sec487.info/p.

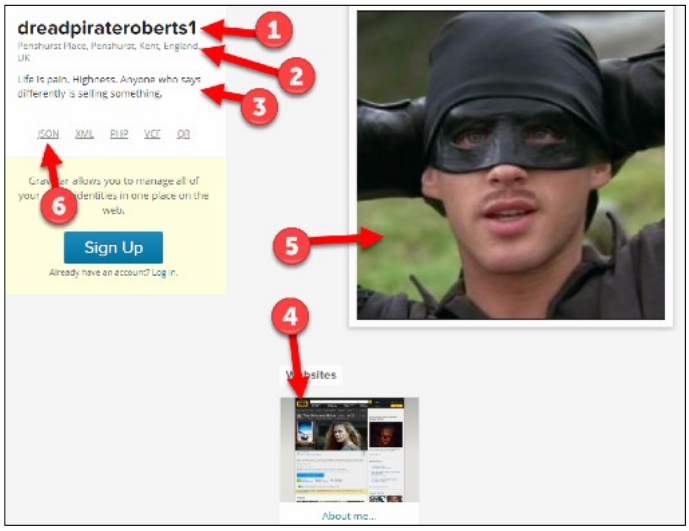
The output from the site defaults to HTML (normal web page view) but can be altered by appending .json, .xml, or .php to the end of the URL request, depending on if you would like to see the output in JSON, XML, or PHP format, respectively.

[1] http://en.gravatar.com/, July 27, 2016.

Gravatar HTML Output

For OSINT purposes, the profile page has many data points to collect:

- User name (1)
- Location (2)
- Interesting quotes (3)
- Web site (4)
- The avatar (5)
- JSON content (6)



Gravatar HTML Output

The default web browser and human-friendly output from Gravatar contains the information that the user has posted to the site. This data can contain user names, locations, target personal data, and, of course, the avatar image itself.

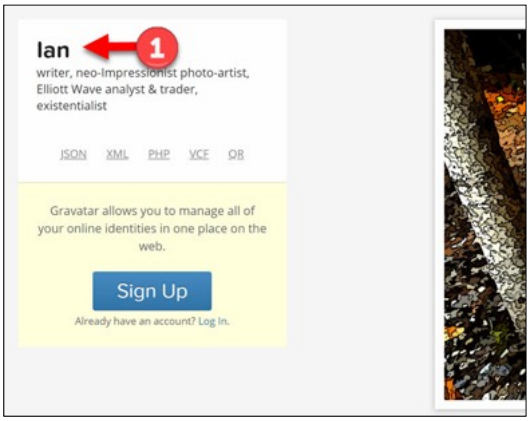
All this content should be recorded by the analyst and examined for meaning. It should also be corroborated against other data collected for possible false positives. Analysts should remember that this information is provided by a target and may or may not be truthful and accurate.

In the slide above, on the right side is an image of the "dreadpirateroberts1" Gravatar profile page (https://sec487.info/q). Data worth harvesting for your OSINT investigation is noted above. Please understand that this is a sample profile and may not have all the fields of a real person's account.

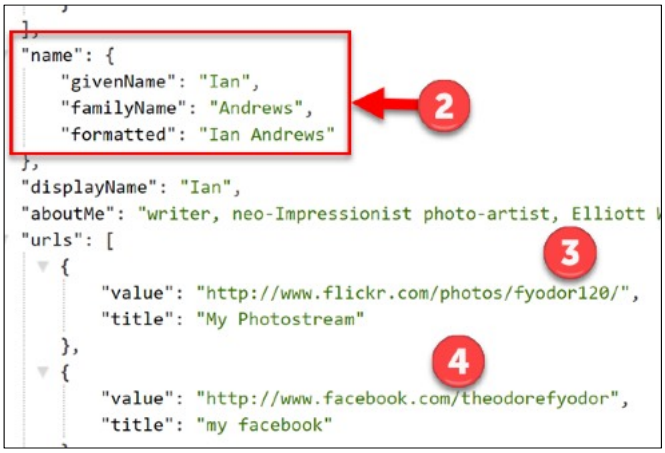
Image from https://sec487.info/q, September 4, 2019.

Gravatar JSON Output Has More Data

gravatar.com/fyodor



gravatar.com/fyodor.json



Gravatar JSON Output Has More Data

The JSON output of the Gravatar site shown above for the user "fyodor" displays some of the content we saw previously on the HTML web page. Note the user's display name of "theodore" (1). What many people do not know is that web applications many times provide additional details not found in the HTML format via their JSON output. If we click the "JSON" link on the web site page on the left, we get the JSON output on the right.

Notice that in the JSON output, we now have additional names that the user submitted to Gravatar.com when they filled out their profile. The data found here will vary based upon what information your target has placed into their profile.

Images from https://sec487.info/rd and https://sec487.info/re, November 9, 2019.

Image Search Engines

- Sometimes called "Image Match" or "Reverse Image Search"
- Image search engines look for an image across all their indexed sites
- Only for unauthenticated web sites that have been crawled
- Ask "Why is this avatar in these results?"

Image Search Engines

There are several image search engines on the Internet that will allow us to submit an image and perform a search for other places on the internet where that image appeared. You can submit an image to be analyzed by uploading the picture from your local computer or, if the image is publicly available on the internet, by providing a URL to the picture.

The search engine will then search for that image across all the sites the engine has indexed. On some sites this process is named a "reverse image search" because you provide the picture of what you would like retrieved to the search engine and it will show you where it found the image and pull up text. Other sites refer to this as an "Image Match" search. Keep in mind that most search engines do not authenticate to web sites when they crawl them. What we are looking for are images on the public internet that match your image.

As OSINT analysts, we ask questions such as "Why did this target show up on the results pages?" Sometimes the answer is simple: the avatar/picture that was submitted was located on other sites where that user may have an account. In this situation, the OSINT analyst should note the responses and systematically visit each site to determine if the avatar matches are valid and indicate another account for the target.

The results also may show the original location of an image that the target took for their avatar. This valuable information is worth noting and exploring.

On the next slide, we will examine one aspect of positive image search results further.

Submitting Images

- Typing in the URL to the image is easiest
- Can upload from local computer on most sites
- Beware that, in some cases, the URL to an image doesn't end in an image file extension
  - For example: GitHub https://avatars3.githubusercontent.com/u/20526634?v=3&s=460
  - In these cases, download the image to your computer and then upload the image to the search engine

NOTE: You may need to crop the image before searching

Submitting Images

There are multiple methods of submitting images to search engines. If the picture is on a publicly facing system, the simplest of these is to copy the URL to the image and paste it in the search engine. This works well when the image file ends in a traditional picture format file extension such as PNG, JPG, or GIF. What about sites such as GitHub.com whose URLs to avatar images look like https://avatars3.githubusercontent.com/u/20526634?v=3&s=460 (https://sec487.info/r) with no traditional filename extension at the end? Some search engines will try to retrieve the image and then process it, but many won't simply because the file lacks the typical filename.extension format.

In these cases, we resort to using the browser's "Save Image As…" feature available in most browsers by right-clicking on an image. After we save the image locally, we can then use the image search engine's "Upload a picture" feature to execute the search.

Note that sometimes the images we wish to examine may have been altered, enhanced, and otherwise modified before we received them. If you are not receiving good search results from these search engines, you may wish to try cropping the photo and uploading that smaller image to the engines. It is suggested that you pick a distinct section of the image to upload (examples might be a close-up of a face or a specific building).

Search Engines for Images

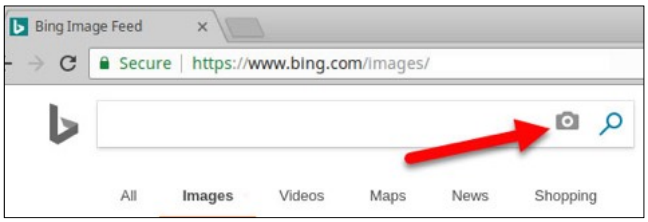
- Many reverse search engines exist
- If we are going to use one, we want the search engine to have a wide sampling of internet images to compare
- Three of the top image search engines are
  - Google Image Search
  - Bing Image Match
  - Yandex Images Search

Search Engines for Images

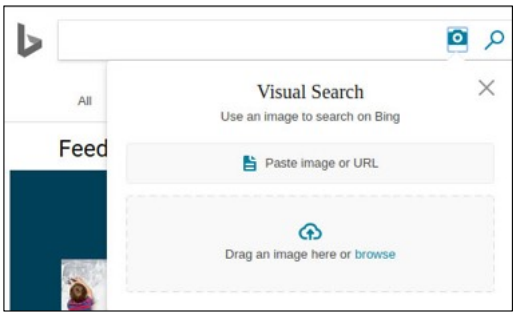
So, you now have either an image URL or a file locally on your system. Now, you as the analyst have a choice for which engine(s) you upload it to. Let us explore three of the most popular sites: Bing, Google, and Yandex. We want the search engine we use to have a wide range of images to compare our image to, and we want the results to be accurate.

Bing Visual Search

Bing.com main images page displays search field



Once the camera icon is clicked, the Visual Search box requests an upload or image URL to process



Bing Visual Search

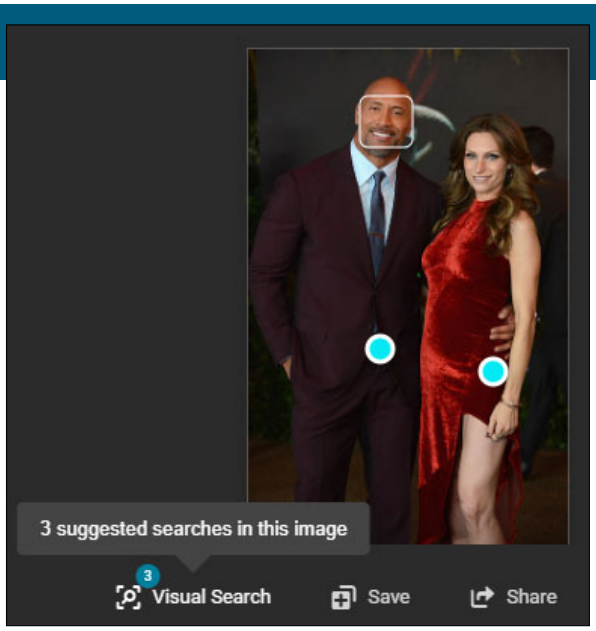
Microsoft's Bing search engine has a competent reverse image search feature called "Visual Search." When visiting the Bing website, ensure you click the "Images" tab under the search box or visit https://sec487.info/s. In the picture above, users can click the camera icon (with the arrow pointing to it) to display the place where they can upload an image or enter a URL where Bing can download an image.

Bing Visual Search Live Cropping

Bing.com image search allows for live cropping of images

This can reveal other images that have portions of your target image

For popular images, they may be able to tell you who the people are and what they are wearing



Bing Visual Search Live Cropping

Sometimes, when you send an image to an image search engine, no results are returned. In these cases, cropping the image and resubmitting can yield some results. Instead of editing the image on your system and uploading it to Bing several times with different portions cropped, you can use the Visual Search feature on your web interface to move the blue corners of a rectangle around and virtually crop your image. As you perform these modifications to the area of the image to be searched, the right side of the web page reveals any image matches…in real time.

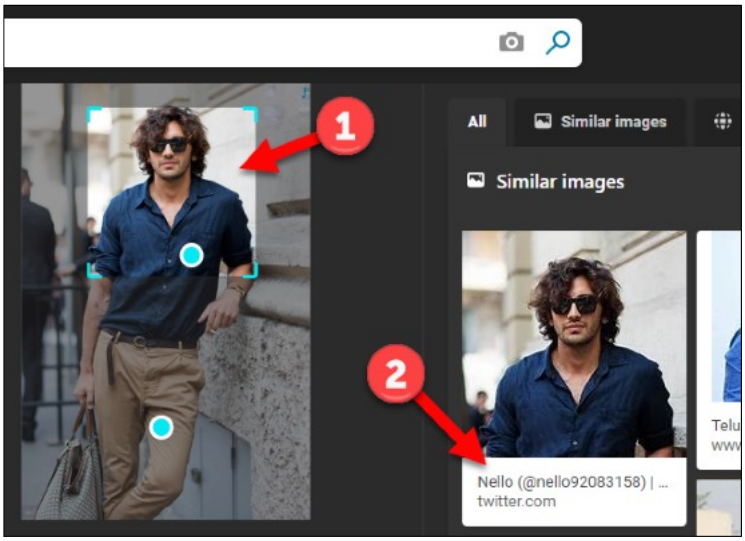
Images may have certain regions highlighted by Bing. Faces, clothing, and other easily recognizable sections might have blue dot "hot spots" that you can click to get more information on those areas of the image. In the image above, clicking the blue dot on the left will search for men's suits, and the one on the right will search for images of red dresses.

Image from https://sec487.info/qb, August 21, 2019.

Bing Visual Search Live Cropping Results

By cropping images using the Visual Search feature, we can focus on portions of the image

In this case, someone used the upper portion of this advertisement (1) for a Twitter account profile image (2)



Bing Visual Search Live Cropping Results

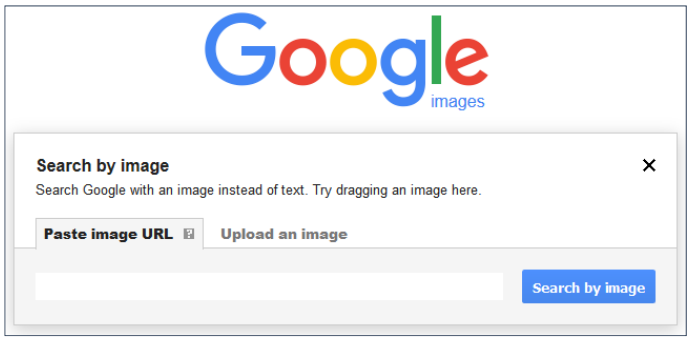
Using the live cropping or visual search on Bing, we can focus the search engine on a portion of the image, such as the upper half of this advertisement. Doing so changes the results on the right to reveal that someone used this portion of the image in a Twitter profile with the handle @nello92083158 (which has been suspended by Twitter).

Image from https://sec487.info/qc, August 21, 2019

Google Image Search

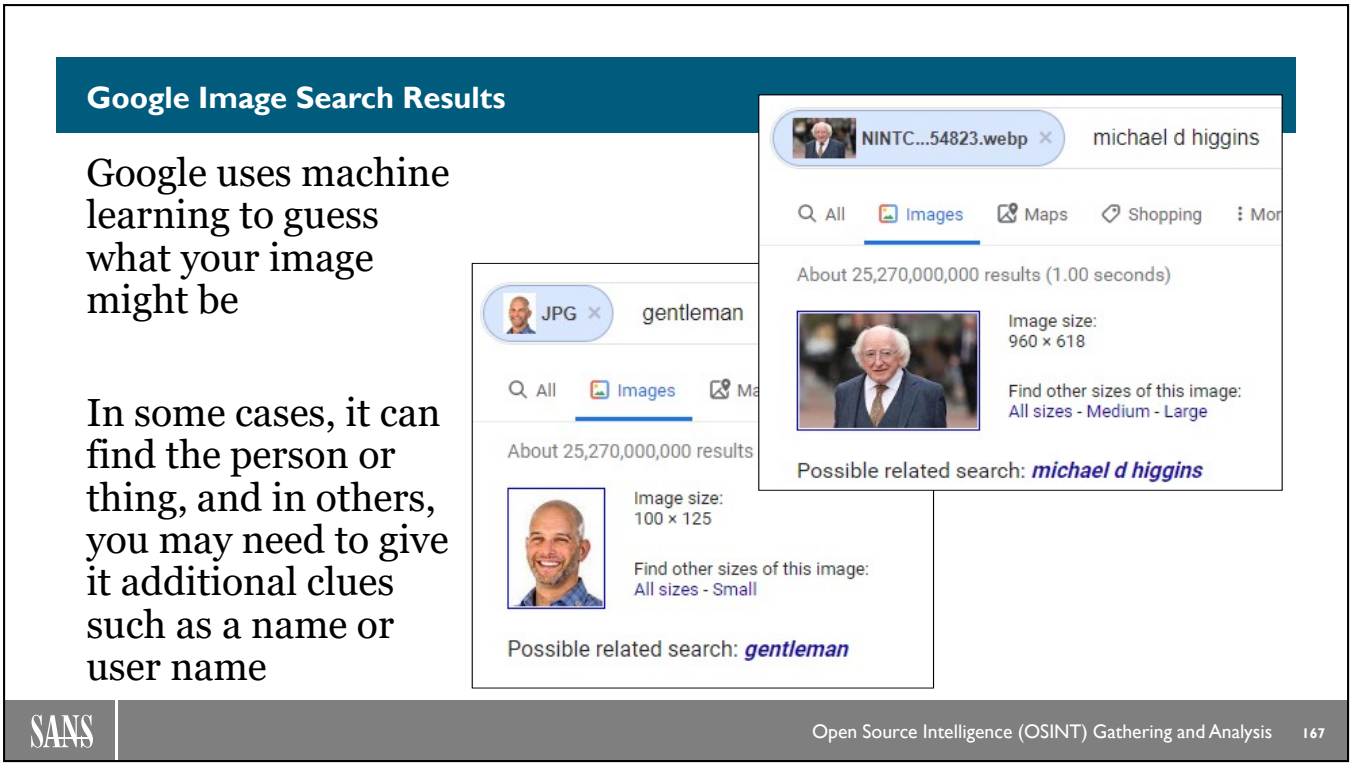
Google image page shows a camera for image search

The secondary box pops up, allowing image upload or URL submission



Google Image Search

Google's image search can be found at https://sec487.info/t. On this page, the search box has an icon of a camera inside it (see the arrow in the first image on this page). When this camera is clicked, another dialog box pops up in the browser. This box has options for the user to drop a picture on the page, upload a file in the traditional manner, or submit a URL to be processed.

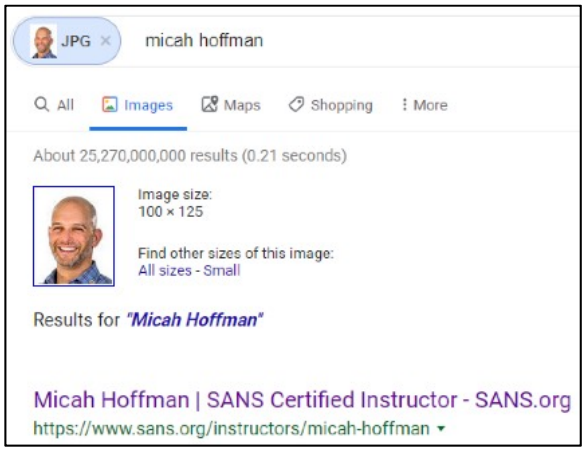


Google Image Search Results

With Google's massive index of web sites, it is a great site to start off an image search. OSINT research is many times about finding as much as possible about a target in the shortest amount of time. Google's image search analyzes the resource at the given URL location and tries to figure out what it is. Give it a picture of a person, and it will try to discern who that person is.

In the slide image above, the 9th President of Ireland, Michael D. Higgins, was simple for Google to recognize (https://sec487.info/yk). But an image of Micah Hoffman from the SANS Institute was not. Google's engine broadened its search and picked a less-specific term like "gentleman" to describe the picture.

We can focus Google on the image being associated with Micah Hoffman by removing the word "gentleman" and replacing it with "Micah Hoffman," as shown below (https://sec487.info/qd).

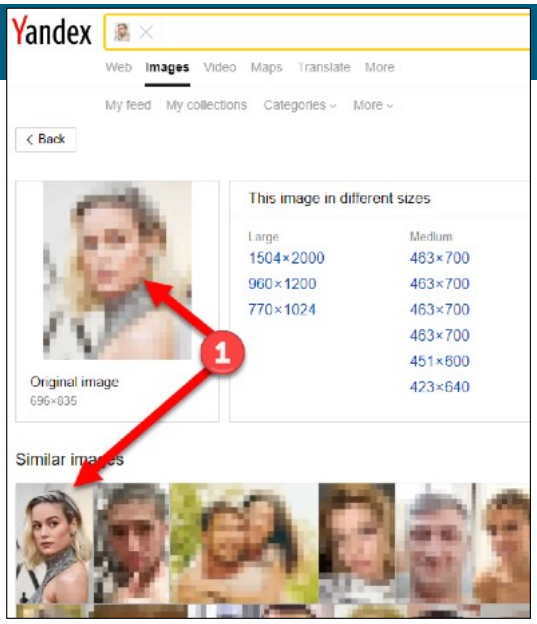


Yandex Image Search

Yandex's image search is excellent at recognizing people, faces, and content from or about Europe and Asia

Click the camera icon to upload an image or paste a URL

The image on the right shows how Yandex can find a matching image even when it has been "blurred"



Yandex Image Search

Yandex.com, a Russian corporation, focuses some of its search engine and resources on Asian and European targets. However, its image-searching features are outstanding when looking for human faces in matching images. Upload your image or point Yandex's image search feature (https://sec487.info/r9) at an image, and it will perform a search. With its focus being on Russian sites, some of the search results may by indexed in Russian or other similar languages.

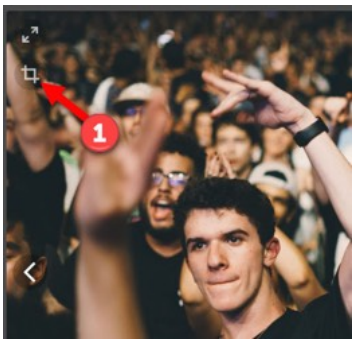
Image from https://sec487.info/qe, August 21, 2019.

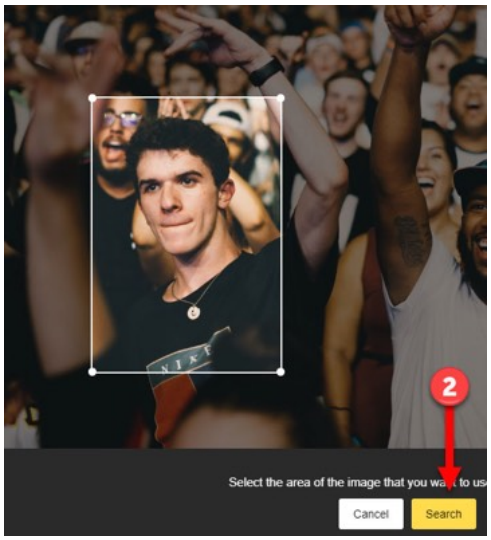
Yandex Live Cropping

Yandex has a live cropping feature similar to Bing

Click the crop icon (1) and then select what you want to search

Then click the Search button (2)





Yandex Live Cropping

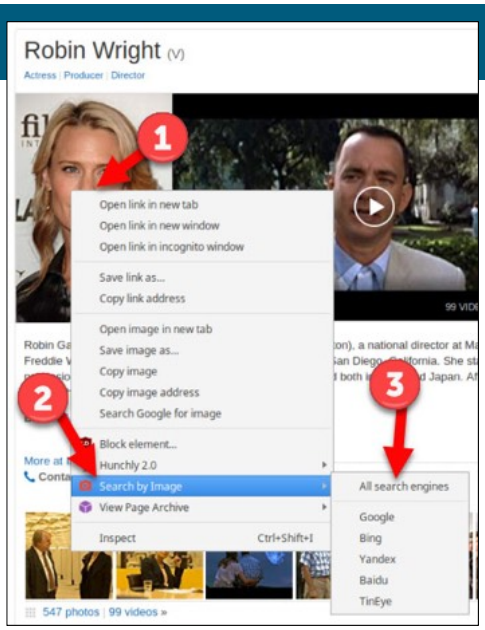
Similar to Bing's "Visual Search," Yandex has a feature where you can perform image cropping within its web page. This makes it easier to search for a portion of an image without having to download it, crop it, save it, and upload back to the internet. In the above image, we click the crop button (1), adjust to the size we want (in this case, highlighting the man's face and necklace), and then perform the subsequent search (2).

Images from https://sec487.info/ra, September 4, 2019.

Easily Searching for Images

We could visit each image search site individually, but it would be time consuming

Adding the "Search by Image" (by Dessant) extension to Firefox or Chrome speeds this up



Easily Searching for Images

Like many processes in OSINT, we could do they manually and we would be very successful. The downside to doing that is that you spend your time performing tasks that could be done in a more automated fashion. This means your attention is pulled away from doing those tasks that require manual review and analysis.

As we have seen in this module, we have several search engines that we need to search to find information about our images. Instead of visiting each, locating the button to provide it a URL, and then submitting the URL, we can use a browser addon or extension to do this work for us. Armin Sebastin's Search by Image extension[1] for Chrome (https://sec487.info/rk) and Firefox (https://sec487.info/rl) does the repetitive work for us. Simple find an image that you wish to search for across one or more search engines, right-click the image (1 above), move down to the Search by Image context menu item (2), and then select which search engine you wish to send the image to (3). The extension will open one or more tabs in your browser for each search engine you asked it to search. The last step is for you to analyze the search engine results for a potential match to your image.

Reference and image: [1] https://sec487.info/ri, September 6, 2019.

Image from https://sec487.info/rj, September 5, 2019.

Analyzing Local Images

Sometimes you will have many image files on your analysis system to review

We have techniques to parse the contents of these files more efficiently than opening each and having an analyst viewing the content

We saw earlier how to parse PDF files for text content

Image files can be examined in a similar fashion using different free tools

Analyzing Local Images

During investigations, analysts may retrieve hundreds of images to their local OSINT computer. These files need to be parsed for content, and having an automated tool give them a quick once-over can speed up the identification of important data.

Tesseract: The Free OCR Tool

Using Object Character Recognition (OCR), we can extract some text from images

Tesseract is a free, cross-platform OCR command-line tool to extract text from images

Might already be installed in your OS and is available on GitHub

Usage:

1. Download an image with text content in it
2. Run Tesseract on it

```
$ tesseract IMG OUTPUT
```

3. Receive text output
4. Analyze

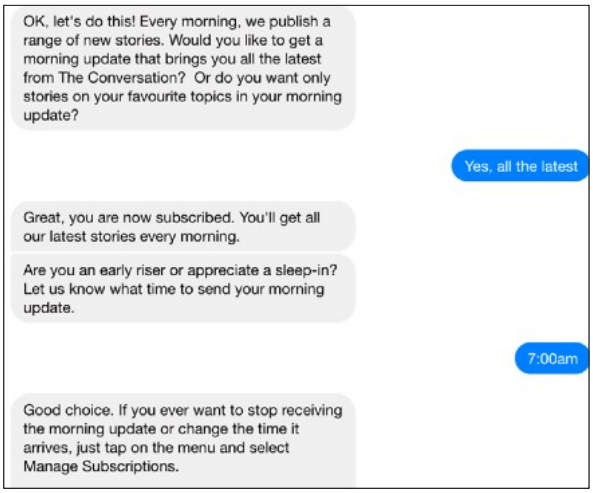
Tesseract: The Free OCR Tool

An excellent, free OCR (Object Character Recognition) tool that works across most operating systems is Tesseract. Download instructions and binaries are available on GitHub (https://sec487.info/lg) for Tesseract and its language-specific packs. When run from a terminal or command line, Tesseract examines the target image file(s) and extracts text from them into a separate output file. That text file can then be analyzed using text-based tools such as grep, findstr, and text editors.

The command to run Tesseract is, in its most basic form, tesseract IMAGENAME OUTPUT (where IMAGENAME is the name of the file that has the content you are looking to analyze and OUTPUT is a filename for the output text or the word "stdout" to show the text output to the screen).

Example Using Tesseract

Facebook Messenger Conversation



Tesseract Output

```
$tesseract fb-msgr.png stdout
OK, let's do this! Every morning, we publish a
range of new stories, Would you like to get a
morning update that brings you all the latest
from The Conversation? Or do you want only
stories on your favourite topics in your morning
update?

Great, you are now subscribed. You'll get all
our latest stories every morning

Are you an early riser or appreciate a sleep-in?
Let us know what time to send your morning
update.

Good choice. If you ever want to stop receiving
the morning update or change the time it
arrives, just tap on the menu and select
Manage Subscriptions.

Yes, all the latest
```


Example Using Tesseract

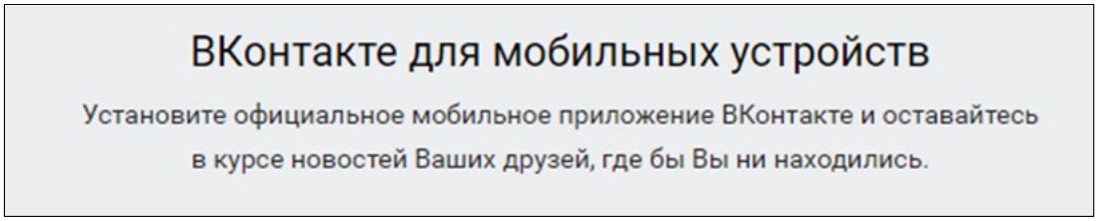
Tesseract works best with images showing clear text. In the left image above, we have a sample Facebook Messenger conversation from the https://sec487.info/lh site. On the right part of the above slide, we see the output text from tesseract. While the text was extracted nearly flawlessly, we can see that the order in which the text appears in the document was altered. Tesseract converted the text from the left side of the document first and then appended some of the content from the right side at the bottom.

This example highlights why it is important to understand how your tools work before you use them in an investigation.

Image from https://sec487.info/lh, December 31, 2018.

Example Using Tesseract Language Pack

```
$ tesseract IMGNAME OUTPUT -l LNG
```




```
$tesseract vk.com.rus.png stdout -l rus
ВКонтакте для мобильных устройств

Ушннпвите официальное мобильное приложение ВКоишкте и
пспвайтесь
в курсе нпвптй Ваших друзей, где бы Вы ни находились.
```


Example Using Tesseract Language Pack

Tesseract's main OCR capability can be augmented with many language packs that help it understand the different character sets in many languages. An example of this is shown above, where we visited the https://vk.com web page and told it to show content in Russian. Taking an image of that content and running it through tesseract and specifying the Russian language using the "`-l rus`" option yields Cyrillic characters in the output.

To use the language packs, you will need to install the ones for languages you are looking to use. The https://sec487.info/lj site on GitHub displays all the language packs that can be used along with their three-character abbreviations. The https://sec487.info/lg page discusses how to install the language packs.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine and begin the exercise named:

"Reversing Images"

This page intentionally left blank.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. **Additional Public Data**
10. Creating Sock Puppets

This page intentionally left blank.

OSINT Value of Registries and Wish Lists

Registries and wish lists give insights into people's wants

They also are links to other places to find personal information and people that are important to your targets

Finding a registry can yield opportunities for social engineering too

General pattern of registry OSINT:

1. Use searches to find registry
2. Examine other data for wedding web site, who bought gifts, etc.
3. Pivot to other places to find more OSINT data
4. Repeat as needed

OSINT Value of Registries and Wish Lists

The obvious answer of why gift registries and wish lists are important in some investigations is that they are usually to celebrate an event. Events have people associated with them, and we are looking for people! So aside from knowing what specific gifts people are desiring for their wedding, birthday, or baby shower, we see who their friends and family are in the comments and in who is buying the gifts.

In general, we use registries as pivot points to find other people, locations, photos, dates of events, and OSINT data of interest. We continue to pivot through the data we retrieve and use it to search social media sites.

Where to Look

There are web sites that allow us to search for wedding registries:

- registryfinder.com
- myregistry.com
- theknot.com
- thebump.com
- Amazon (https://sec487.info/ae)
- honeyfund.com (Virtual honeymoon)

Where to Look

There are several wedding site or wedding registry aggregators. These web pages, shown above, allow users to search databases for the couple that is registered. Once located, there are usually links to other sites for registries and wedding web sites. We can follow those links to gain more details about our targets.

Wedding registries are very common in countries around the world. But weddings are not the only times people submit wish lists for gifts. We can find wish lists and registries for birthdays, anniversaries, graduation from schools, and baby showers. Many online retailers (Walmart, Target, Macy's, etc.) have registries that can be searched.

Image from https://sec487.info/a7, November 4, 2017.

Wedding Web Sites

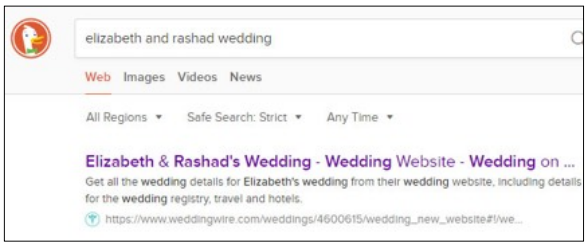
Another place where people share a large amount of data is on wedding web sites

When some people get married, they create a web site about the event and share it

The site can have images, videos, links to registries, location/dates, guestbooks, and bios

There are many web sites that host wedding pages

We can use a search engine and look for key terms such as "person and person wedding"



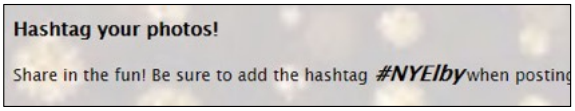
Wedding Web Sites

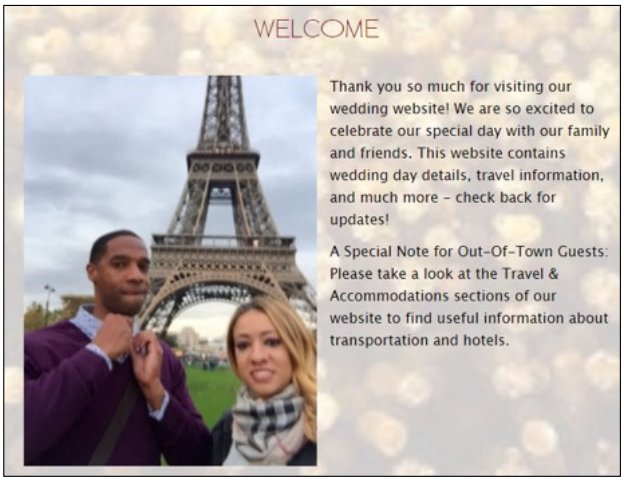
A corollary of the wedding registries is that some couples, when they decide to get married, create a wedding web site. These sites, often hosted at the same places that their registries are registered, contain a wealth of data that OSINT analysts can obtain and pivot upon. Among the data we can find on these sites are images, videos, links to registries, people, and biographies of the couple.

If you know the couple's first names, you can use a search engine to search for "name & name wedding" and many times the wedding web page will be in the results. In the above image, we searched for "elizabeth and rashad wedding". Let's take a look at the first result and use this as an example.

Image from https://sec487.info/a8, November 4, 2017.

Elizabeth and Rashad on WeddingWire





Elizabeth and Rashad on WeddingWire

The WeddingWire web site hosted the wedding pages of Elizabeth and Rashad (no last names for the couple appear on the site). Right off, we have the date of the wedding, December 31, 2016, and pictures of the couple. We also have a hashtag of #nyelby that can be used on social media.

There are other links on the page that show the wedding event was held in Indianapolis, Indiana. Photos of the couple are also shown.

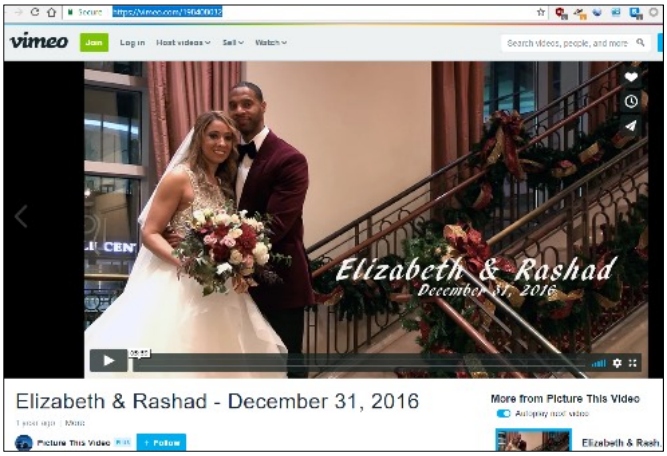
Images from https://sec487.info/a9, November 4, 2017.

Pictures and Videos from Google Search

Videographer's Blog

Congratulations to Elizabeth & Rashad, who were
Year's Celebration at the Hilton Hotel, http://www3.hilt
suites-INDDNHF/index.html, in DT Indy, . Elizabeth
Hilton with hair and make-up, by Prestige Salon and
the groomsmen met and prepared for the day on the
first look with each other in the ballroom where the c
Curry Photography, took shots of the first look and p
ceremony more photos were taken while guests enjoy
and of course us with video walked to the Conrad
restaurant "Tastings" where Elizabeth & Rashad ac
provided by First Impression Band, www.firstimpres
Champaign Toast complete with countdown at midnig
ready events. LLC for planning and organizing the da
great job of making the day run smoothly for us. Cong

Vimeo Video



SANS Open Source Intelligence (OSINT) Gathering and Analysis 181

Pictures and Videos from Google Search

Performing a Google search on the couple showed results for the www.picthisvideo.com site, which hosted information about the event, photos of the bride and groom and their guests, and a video hosted at the Vimeo web site (right picture above and no longer available online).

Images from https://sec487.info/hn, March 31, 2018.

Video Details

Vimeo video shows copy of the marriage certificate with date, time, and county (among other things)

Video Details

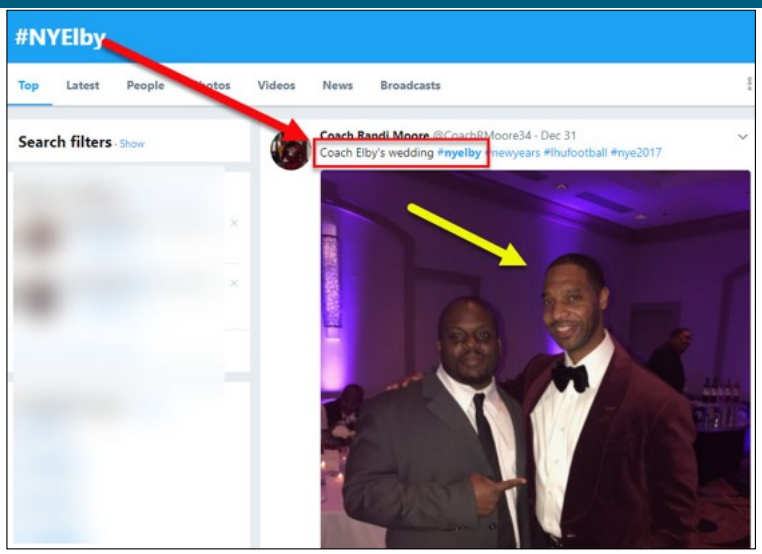
The video for this event still does not show last names for the bride and groom but does confirm that they were married in Marion county, Indiana on the 31st of December 2016. To pull their completed marriage license from the Indiana government pages, we still need their last names.

Back to the Hashtag - #NYElby

Heading over to Twitter and searching on the hashtag:

#NYElby

We get hit and see: "Coach Elby" and a picture of him



Back to the Hashtag - #NYElby

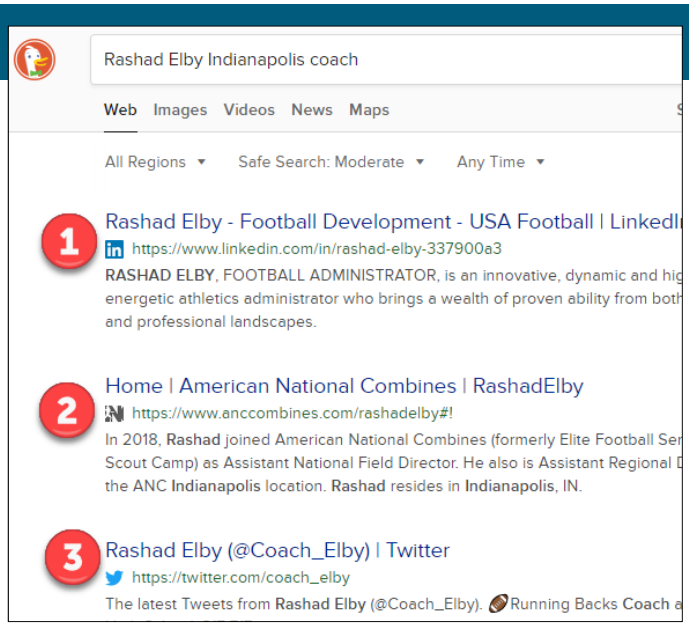
Sometimes we need to go back to the beginning and pivot on another bit of data to find what we are looking for. So, let's go back to the hashtag that we saw on the wedding page: #NYElby. When we try it on Twitter, we get one result (shown above). The Twitter user tweeting this post mentioned the hashtag and "Coach Elby's wedding." Could "Elby" be the groom's last name or is that a nickname?

Image from https://sec487.info/aa, November 4, 2017.

DuckDuckGo Search

Since we know the first name (Rashad) and possible last name (Elby), perform a DuckDuckGo search for:

```
Rashad Elby
Indianapolis
coach
```



DuckDuckGo Search

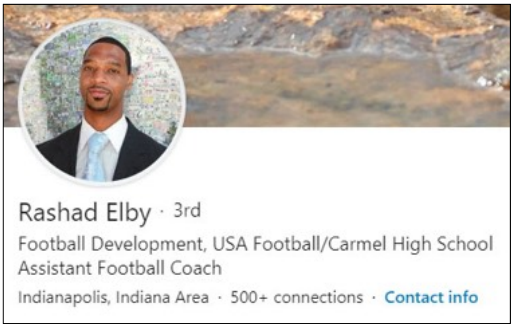
Now we have a possible last name, Elby, to go along with the first name Rashad. With this information, his title "coach," and the location we found earlier (Indianapolis), we construct a search query. Using DuckDuckGo.com, we search for "Rashad Elby Indianapolis coach" (https://sec487.info/gu) and we get several results, including a potential LinkedIn and Twitter profile.

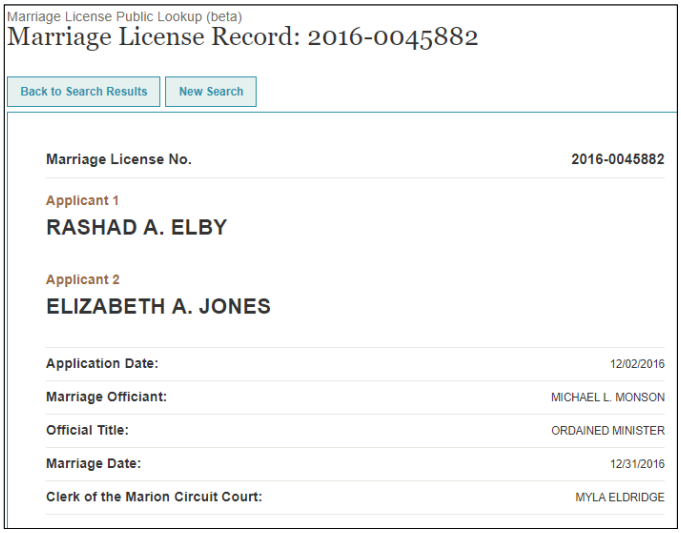
The results of our search yield other web sites to visit and verify that this is our target.

Image from https://sec487.info/gu, October 5, 2019.

LinkedIn Profile Confirmation and Marriage License

We get his LinkedIn profile with full name then search Indiana marriage licenses





LinkedIn Profile Confirmation and Marriage License

Visiting the LinkedIn page (as an authenticated user) we found, we can see that the picture, first name, and location all match for our target. His last name is Elby. Entering that information into the Indiana state government marriage license search brings up his license with the full names of himself and his bride, Elizabeth A. Jones.

Images from:

https://sec487.info/ac, October 5, 2019.

https://sec487.info/ad, November 4, 2017.

Obituaries and Death Notices

Death notices are published in newspapers and online, so people can show support for the family of the deceased

Many of them provide a vast amount of family data, date of birth/death, photo and location data

Sensitive information such as US Social Security numbers becomes public data at death

OSINT use:

Search for a family member of your target and see if other familial data is exposed

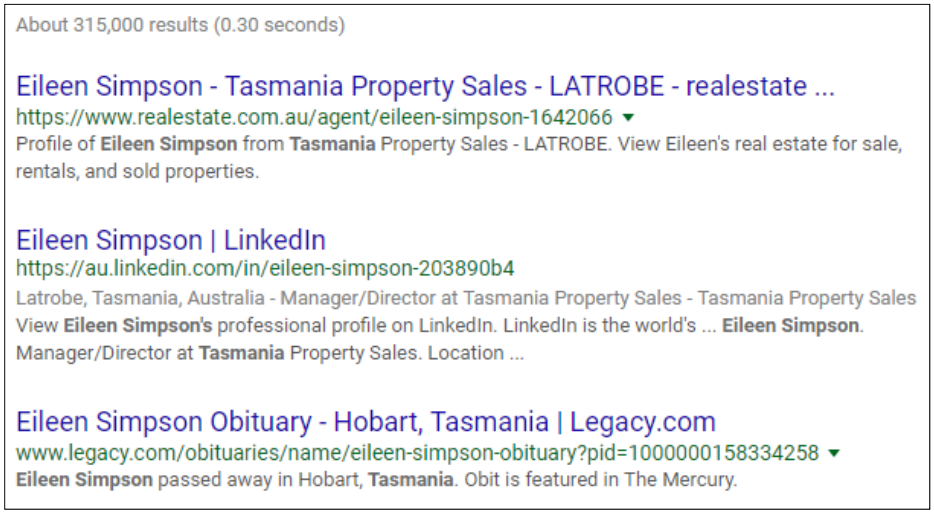
Obituaries and Death Notices

When a person dies, their family may choose to create a death notice or obituary about their loved one. These paragraphs usually note the person that died, where and when the passed, who they are survived by, things they enjoyed doing, occupations, and sometimes a photo. These obituaries can be published in newspapers and in online venues such as funeral home and online publications. You can see that these notices contain valuable OSINT information on the families of the deceased. If your target is someone who has died, or you need data about their family members, obituaries are going to be a great source of information.

For our OSINT purposes, search for either your target (if they have died) or for a deceased family member of theirs and try to locate a relevant obituary. Harvest the data from the obituary and pivot on it as needed.

Google Makes It Easy

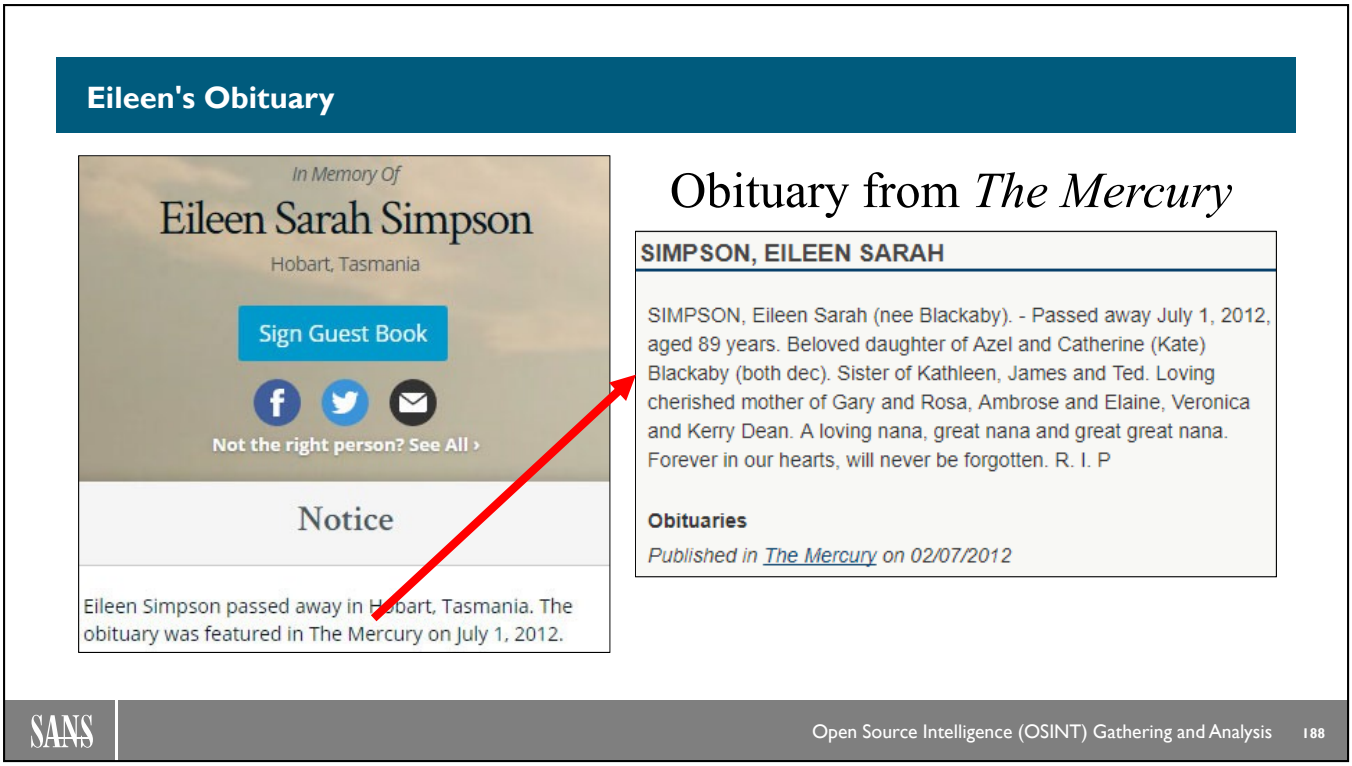
Here is an example of searching for a realtor in Australia



Google Makes It Easy

Here is a quick example of how simple search engines sometimes make our work. Let's say that we are interested in finding more information about a specific real estate agent in Tasmania, Australia. Her name is Eileen Simpson. The more public a person's life (actresses/actors, politicians, sales people, etc.), the more information about them can be found online. Entering the search terms "Eileen Simpson Tasmania" into Google's search engine (https://sec487.info/4n) shows the top three results in the slide above. The first entry is for her office/work web site. The second shows a link to her social media profile on the LinkedIn web site. And the third result shows that she died. Let's continue with this OSINT example for another moment.

Searching for death notices in search engines is simple. Enter `"death notice"|obit` and your target's name into the search field and execute the search.



Eileen's Obituary

Continuing with the OSINT portion of this example, let us say that your OSINT work specified that you needed to find the names of Ms. Simpson's family members. Perhaps your client had found an old insurance policy but didn't know who to send the money to. Following the link to the obituary page (https://sec487.info/4o), we get her middle name (Sarah), where she died (Hobart, Tasmania), that her obituary was in *The Mercury* (most likely a local newspaper), and it was posted on July 1, 2012. A few clicks and searches later, and we can retrieve that obituary from *The Mercury*'s web site (https://sec487.info/4p and shown above).

With the full obituary, we have the names of her entire immediate family, including parents, siblings, and children. Next steps? Perform searches on those family members and give them the insurance policy money!

Legacy.com for Death Notices

We can also use a dedicated death notice web site such as legacy.com to search for our targets

It not only contains some obituaries but also links to funeral homes and newspapers for you to get additional information

Obituaries for the following countries are supported:

- USA
- UK
- Ireland
- Australia
- Bermuda
- New Zealand
- Canada

Legacy.com for Death Notices

The legacy.com web site contains a wealth of information about people who have passed away in some English-speaking countries. In addition to containing some excerpts of actual obituaries, it can refer you to other sites that may have relevant data.

Additional Memorial Sites

Look for memorials:

- Where your target lived
- Where your target's family lived
- Where your target died
- In local newspapers
- In religious newsletters
- In community newsletters

Other sites to check:

- ForeverMissed.com
- iLasting.com
- Imorial.com
- Mem.com
- Qeepr.com
- Remembered.com
- Tributes.com
- YourTribute.com

Additional Memorial Sites

Depending upon where your targets lived, where their families lived, and where they died, they may have memorial listings on other sites across the internet. Here we present other sites that you can use to find the data you need. Thanks to https://sec487.info/ko for the references.

Additionally, be sure to check newspapers, religious organizations, and community newsletters around the target's area for write-ups.

US Government Sources for Death Records

The National Archives Death Records Death Files (1936–2007)

Search for a person's name and see what files they might be in

Results per page 10

View Record	SOCIAL SECURITY NUMBER	FIRST NAME	LAST NAME	DATE OF BIRTH (MONTH)	DATE OF BIRTH (DAY)	DATE OF BIRTH (YEAR)
	701103324	SMITH	JAMES	August	22	1902
	340269832	JAMES	H SMITH	June	25	1934
	433185381	SMITH	JAMES	August	8	1916
	449566375	SMITH	JAMES	September	7	1938

US Government Sources for Death Records

The National Archives warehouses a huge amount of data in its systems. There are death records in those entries for people who died from 1936 to 2007. Simply enter the person's name into the search field and press Enter, and all the records with your target's name will be shown.

What data can we get here? The death notices/records contain dates of birth and death, Social Security numbers, full names, and residence postal codes. But there is much more data in these results that can be interesting, including passenger manifests, enlistment records, manifests from ships, and financial reports. Visit https://sec487.info/gx to conduct your own query similar to the one we performed above.

Newsletters

Excellent data in community and religious newsletters:

- Names, Addresses, Phones, Emails
- Who is sick
- Who celebrated a life event and when
- Photos
- Upcoming events

Newsletters

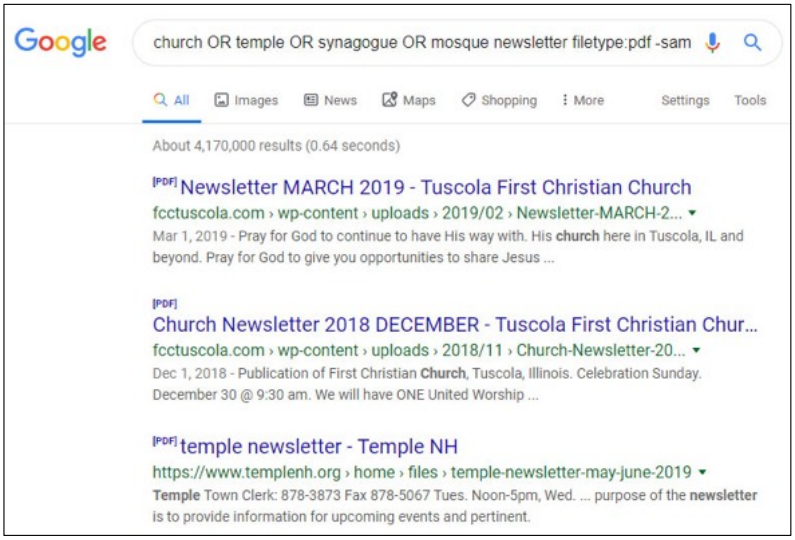
Around the world, people belong to religious congregations, and some of these may publish regular newsletters to the internet. We can harvest photos, personal information, and event data from these documents. They can be a wealth of information, especially if the organization keeps older versions of the newsletters online.

Image from https://sec487.info/kq, June 30, 2018.

Google Knows about Newsletters

August 2019, Google returned over 4.1 million results for:

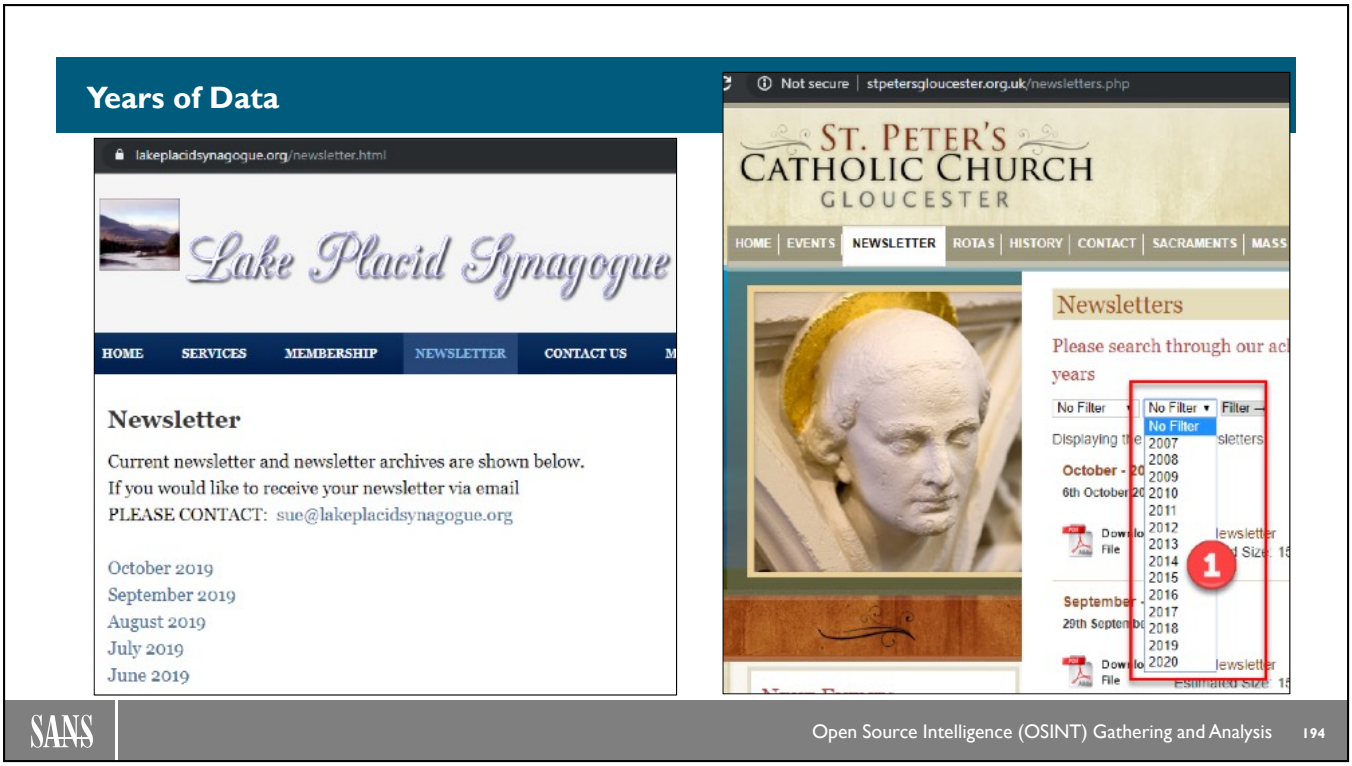
```
church OR temple
OR synagogue OR
mosque newsletter
filetype:pdf
-sample -example
-template
```



Google Knows about Newsletters

Search engines have already found newsletter documents across thousands of web site and allow us to just ask for them. Performing a Google search[1] for "`church OR temple OR synagogue OR mosque newsletter filetype:pdf -sample -example -template`" yielded over 4.1 million results in August 2019. Keep in mind that you can add the `filetype:docx` to get even more documents!

Reference and image from https://sec487.info/qj, August 23, 2019.



Years of Data

While some religious organizations only make available the most recent newsletter, some allow users to downloaded them from previous years. The two shown in the above slide have archives dating back over 5 years with the church going back over 13! When we think about the life events in these documents, in 13 years, a couple could have married, had or adopted a child, and had a religious ceremony for the child...all of which could be captured within these files.

Images from https://sec487.info/kr and https://sec487.info/ks, October 5, 2019.

Example Content in Newsletters

Not secure | www.monflanquin.fr/file/Monflanquin-newsletter-text-in-E

Civil List

14th April - 30th June 2015

Births:

3rd May	BIENAIME Noah, Rakshan
5th May	VALANDE PUCCI Alessandro, Roberto, Giovanni
9th May	TEYSSIER Clémence
2nd June	FIOL SANCHEZ Louane, Jo
4th June	MAUPOUET SCHRAMME Axel, Charles

Marriages:

2nd May	COURQUET Didier, Patrick and THAPTHIMKAEW Phapassi
17th June	ROBERTS Ian, Martyn and LEIGH Janet

Deaths

1st May	MATHIEU Louise, Madeleine, married GOMIS.
6th May	COPIN Andréa, Mathilde, widow BOICHOT
6th May	RENARD Marguerite, Marie, Louise, widow BAILLES
12th May	STEVENS Ross, Harvey
26th May	NEUVILLE Jean, Alphonse, Gilbert
27th May	VIONNET Pierre, Julien, Emile
30th May	PISSON Jean-Claude, René
8th June	BORD Jeanne, Marguerite, Eliette, widow VIDAL.
19th June	HESPEL Jacqueline, Alfréda

Mike and Melissa Walsh
Collin, Aidan, and Charlotte "Charley" Walsh
2049 Edgewood Avenue
Burlington, NC 27215
336-570-8452

Manuel Zelee Tarwo
Baptized: February 4, 2018
Parents: Marcus & Gormah Zelee
Sponsor: Martha Garwo

On THE Move

Address Update
Ben Pudewell
952-988-8852
Beacon Hill - Commons
5300 Beacon Hill Road #107
Minnetonka, MN 55345

- **Ted & Bev Dean** 763-533-0871
- **Jodi Krohn**, 763-234-6634 jlynnKrohn@yahoo.com
- **Renee Lach**, 612-201-8621 tonyreneel@aol.com
- **Roxe Lam**, 763-498-8916 Roxyrox21@gmail.com
- **Rita Lange**, 763-560-1150 lngflowers@aol.com
- **Ele Sherrard** 763-533-9641
- **Anita Umland** 763-537-3485 anitaotsego@aol.com
- **Mavis Vacek** 763-424-5544
- **Ann Witry** 612-432-5586 anniebuddie2@gmail.com

Example Content in Newsletters

Some religious newsletters (as illustrated above) include images, names, addresses, phone numbers, and email addresses of congregants. These files can be downloaded and scraped for data in minutes, yielding an excellent list of contact information that can be used in social engineering attacks.

Images from:

March 2018 Newsletter at https://sec487.info/kt, March 29, 2018. (URL no longer active)

https://sec487.info/ku, March 29, 2018.

https://sec487.info/rb, September 4, 2019.

Data Aggregators

Some companies serve as data aggregators; they combine and normalize a variety of pieces of data

melissa Global Intelligence is one that allows users to view some of its stored information

Their "lookups" page[1] has a variety of excellent forms to help OSINT analysts

3 main levels of access:

- Anonymous/guest
- Registered free user
- Paid customer

Data Aggregators

In the song "Every Breath You Take" from The Police, there are the following lyrics (https://en.wikipedia.org/wiki/Every_Breath_You_Take, https://sec487.info/4s):

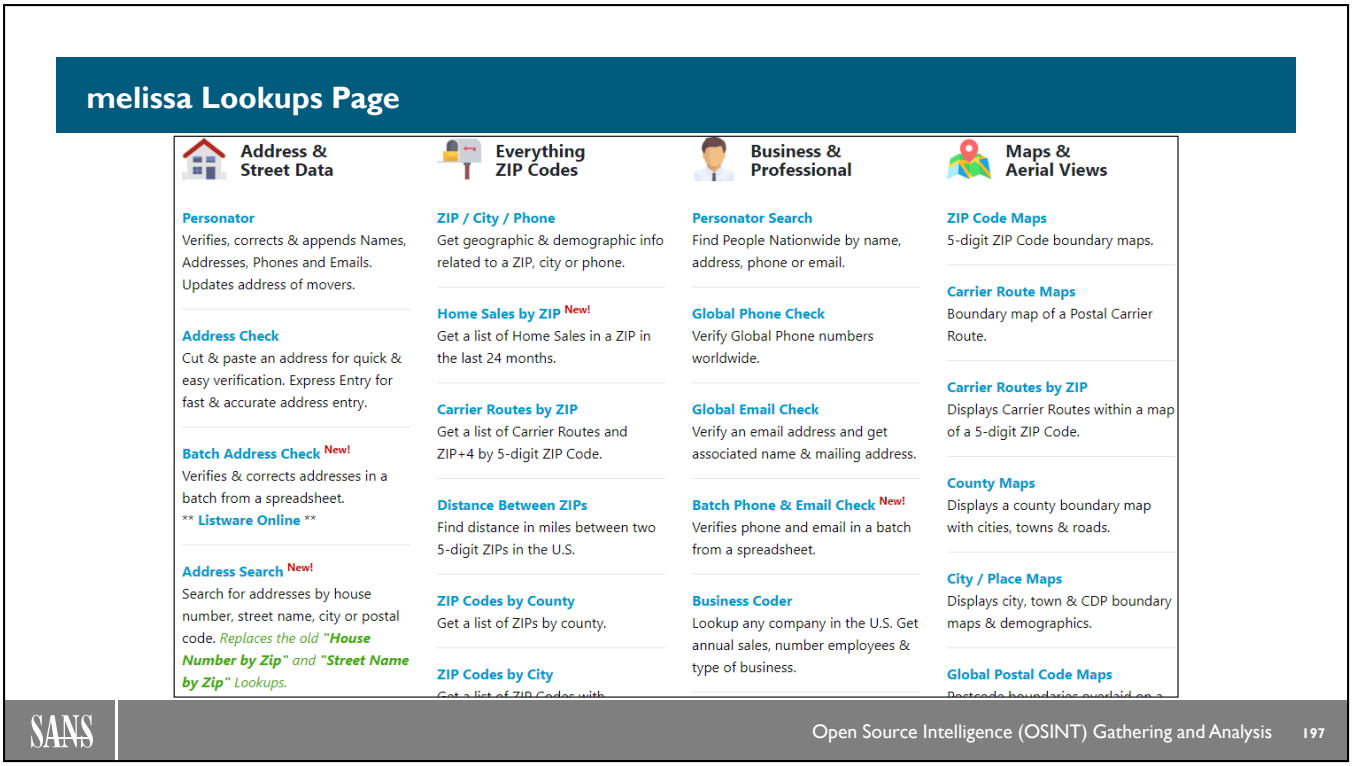
> Every single day
> Every word you say
> Every game you play
> Every night you stay
> I'll be watching you

While Sting wrote this song after he separated from his wife and at the beginning of a relationship with his lover, we can also apply these words to what many companies and governments in our societies are doing to us every day. They collect our words (Alexa, Google Home), where we live, what we buy, where we drive ... all of this information about us is stored. Much of it is beyond the access for normal consumers. Data aggregators, such as Melissa Data, sometimes allow us access to bits of the data that they possess. They get what we buy in stores from credit cards or from customer loyalty programs. They download voter databases (which are public in the United States). They take all this data and normalize it, tag it to people or homes or businesses, and continue to augment it. Sounds devious, yes? Nah, it is just business.

melissa Global Intelligence is one such data aggregator that has tools we can use on their "lookups" page.[1] There are three different levels of access to data on the site: anonymous or guest level, registered user level, and customer level (where you have purchased access).

Reference:
[1] https://sec487.info/4q

Image from https://sec487.info/kv, July 2, 2017.



melissa Lookups Page

There are too many great tools on the melissa Lookups page (https://sec487.info/4q) for us to dive into in depth, and you will enjoy investigating them yourself. As mentioned earlier, some tools require you to register on the site (free) and provide a valid email (they will send you a confirmation to ensure that the email is actually a real one). What do the lookups do? Here is a brief look at some of them that make be helpful in your work:

- **Property Viewer** - Allows the user to select a United States ZIP code and, via a Bing map, zoom into a location. Click on the property you want more information about and melissa will show you who owns the property, market value of the home, how big it is, and the phone number and email of the owner (if it has the data). (https://sec487.info/4u)
- **Global Address Check** - Verify an international address using this tool. (https://sec487.info/4t). There is also one of these just for the United States (https://sec487.info/4v)
- **Personator** - Looks up a person, address, or company. (https://sec487.info/4w)
- **People Finder** - Retrieves more information about a person in the United States. This requires free registration to see detailed data results. (https://sec487.info/4x)

Image from https://sec487.info/4q, October 5, 2019.

Course Roadmap

- Day 1: Foundations of OSINT
- **Day 2: Gathering, Searching, and Analyzing OSINT**
- Day 3: Social Media, Geolocation, and Imagery
- Day 4: Networks, Government, and Business
- Day 5: The Dark Web, Breach Data, and International Issues
- Day 6: Capstone: Capture (and Present) the Flags

GATHERING, SEARCHING, AND ANALYZING OSINT

1. Data Analysis Challenges
2. Harvesting Web Data
3. File Metadata Analysis
4. OSINT Frameworks
5. Basic Data: Addresses and Phone Numbers
6. Basic Data: Email Addresses
7. User Names
8. Avatars and Reverse Image Searches
9. Additional Public Data
10. **Creating Sock Puppets**

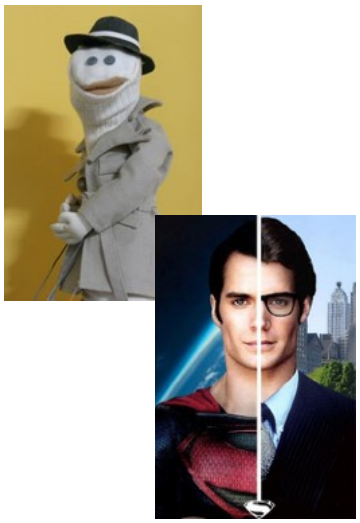
This page intentionally left blank.

What Is a Sock Puppet?

We need to examine content inside sites that restrict access to authenticated users

We do NOT want to use our personal accounts for this access

Sock puppets, research accounts, and synthetic identities all describe false personas we may use on these sites



What Is a Sock Puppet?

While performing our investigations, we will come across certain web sites and APIs (Application Programming Interfaces) that require us to authenticate to the target site to access our target's data. We could use our personal accounts on these systems, but that would be poor operational security (OPSEC) and could notify our target that they are the subject of an investigation and we are the ones performing the work.

Instead of using our own personal accounts during an assessment, we can create false personas, or "sock puppets" (sometimes referred to as synthetic identities and research accounts), that we control for this work. That way, when a site tells our target that someone looked them up, it will give the name of our sock puppet account and not our own. Our sock puppets will be constructed to blend in and look like a normal user of a site to not cause concern in our target.

Images from https://sec487.info/jx and https://sec487.info/jy, August 17, 2016.

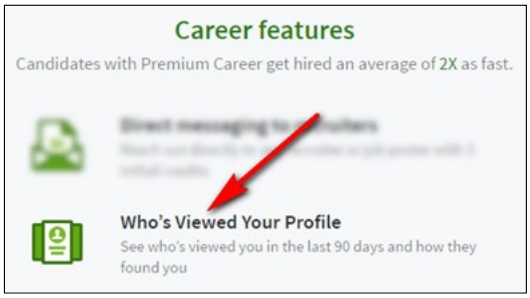
Why Use Sock Puppets?

They create barriers between you and your targets and you and the web application

Social media sites can tell target who you are

- For weeks after an assessment, a site suggested I connect my sock puppet to people with my target's name

From LinkedIn.com



Why Use Sock Puppets?

Since we cannot be sure that information about our web site session is not being shared with our subject, we need to ensure that that information, if it was shared, would not raise concern. Additionally, it may be important to perform our work so that the web site we are using to gather our data does not know who we are and who we are working for.

In some cases, such as with the social media site LinkedIn, shown in the picture above, users can pay the site to show them who is viewing their profiles. Perform an OSINT investigation from your personal LinkedIn account, however, and the target can perform OSINT on you! Finally, some web sites look at who you are searching for and will suggest that you connect/friend/link to those people. For weeks after an OSINT assessment, one of my sock puppets kept getting suggestions from the social media site to connect with people who had the name of the target of the investigation.

Image from https://sec487.info/e0, January 7, 2017.

Acting as Your Sock Puppet

For longer-term OSINT operations, you may need to "take on the role" of your sock puppet

How would this fake person act, think, and feel?

This is usually done when interacting with others on forums, chat rooms, and social media

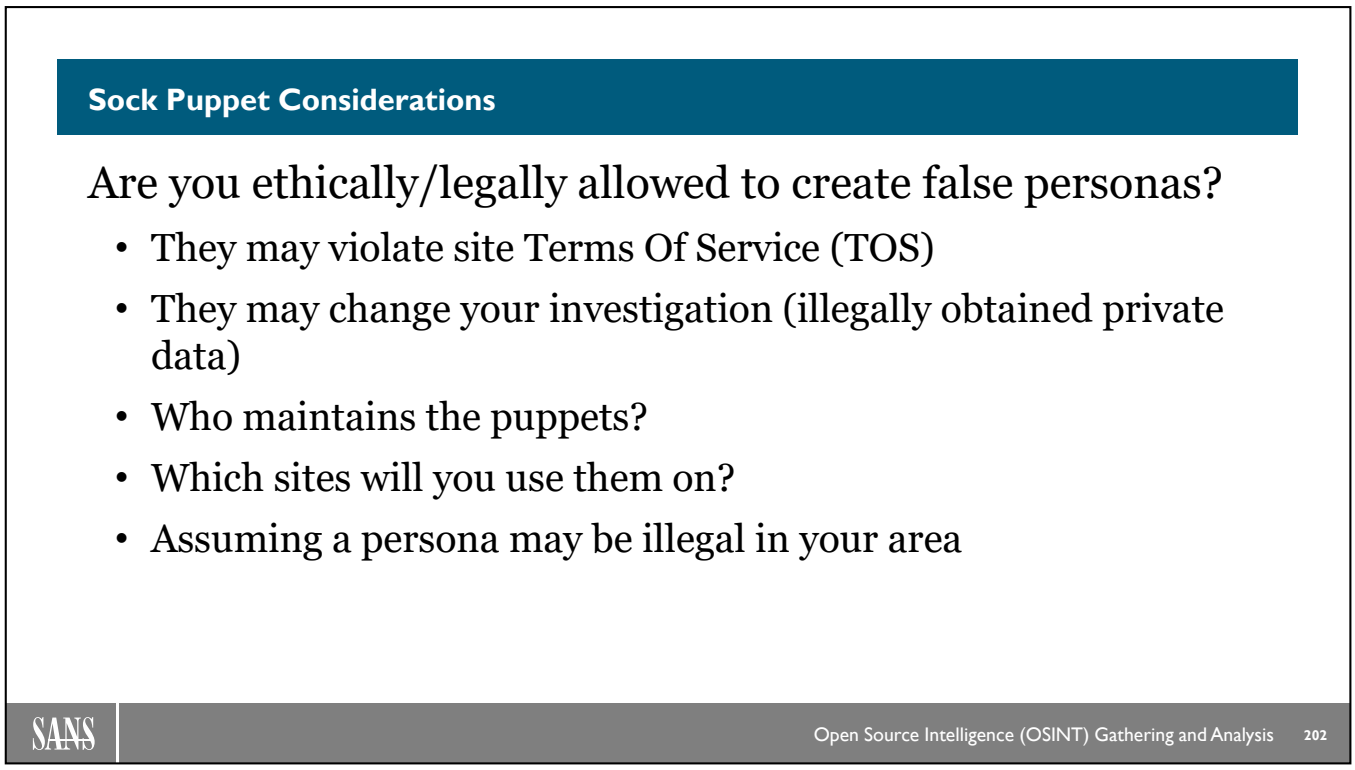
Depending upon the OSINT work you perform, you may create a dossier on your sock puppets, noting behaviors, likes, and accounts

Acting as Your Sock Puppet

In addition to using your false persona(s) to gain access to sites that require authentication, some OSINT analysts will use these accounts to interact with others in chat rooms, games, forums, and social media. In these cases, you, the "puppeteer" of the sock puppet, will need to act as your false persona would. How would a person with the background you created communicate? What would they post? It may be quite different from what you would post on forums and in your social media. Staying in character when using sock puppets can be the difference between a puppet that is taken seriously and one that is dismissed as false.

Working with multiple sock puppets? Have some complex personas that you are working into the fake accounts? You might want to create dossiers of the sock puppets showing biographical data, what they like and dislike, who are their friends, what are their points of view on various topics, and issues like that. This can help keep you "in character" during the assessment.

Of course, if you are not interacting with a target, you often do not need to go through this in-depth sock puppet creation. Just create accounts and make them look "human," and you can begin your assessments.



Sock Puppet Considerations

Creating and maintaining sock puppets can be simple or quite challenging. Without getting into the "how do I make a sock puppet?" question, which we will address later in the course, these false accounts raise certain questions that your team and company need to discuss before you use them.

- Sock puppet accounts are not real people. Because of this, and because some sites specify in their Terms Of Service (TOS) that users of the site cannot provide false data in their profiles, by creating and using sock puppets you may be **violating the TOS of the web sites you visit in your assessments**. For some organizations, customers, and researchers, this is not an issue. However, you need to raise this issue with your legal counsel to understand how this could impact the outcomes of your work.
- If your investigation reports are destined for use in the judicial/court systems, you may not be able to use sock puppets, as you may be **obtaining information under false pretenses**. Again, check with your company's legal counsel prior to using sock puppets.
- We also need to consider the maintenance of sock puppets. **Who on your team will be responsible for making these accounts look like real people**, interacting with others, playing games online, and liking other people's content? It is easy to spot a fake profile on some sites. Look for an account that connects to many other accounts but has no "friends." This is just one of the methods of finding sock puppets or bot accounts.
- **Which sites will you create and maintain sock puppets on** versus surf the data anonymously?
- **Assuming some identities** may be illegal in your area.

How to Make a Sock Puppet

Two methods of making a sock puppet:

1. Only filling in the **essential** information on select sites
2. Creating a **full biography**, a profile, other information, and ties about this fake person to make them appear to be a valid human

You may use a combination of these techniques

One takes more work to create and maintain, but we have some excellent helper web sites

How to Make a Sock Puppet

When thinking about making a sock puppet, you will want to evaluate if you want to create a full persona with activities, emails, phone numbers, social media web sites, and other personal information or if you just want to create an account with the most basic data. The first type of account is useful if your goal is to make your sock puppet appear to humans (and automated anti-bot tools) as a normal human. Having connections, friends, likes, and activities make a profile appear to be used. Alternatively, you may want to only submit the essential information that a web site needs to create an account. We may wish to use this technique, as it requires less maintenance than the complete persona approach and it may still get us the access to our target data. One downside is that these minimalist accounts can appear to be sock puppets to humans and web sites and may get banned or suspended if that violates the Terms Of Service of the site.

The reality is that we will probably employ some of each type of sock puppet. Perhaps on well-known social media networks a fully fleshed-out sock puppet persona is worthwhile, and for more niche sites we might use a minimalist one.

Creating a full persona sock puppet can be time consuming. Think about all the information you need to create: name, address, phone number, email, history, employer, and more! Luckily, there is a free web site that helps us with this.

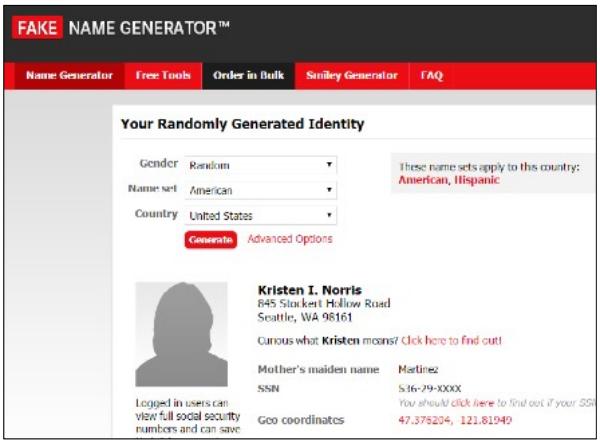
FakeNameGenerator.com and Other Sites

Free site that creates a persona based upon your input.

Returned profile has:

- Name, address, phone
- Family names
- Credit card number too!

The "Advanced Options" page is a great place to start.



FakeNameGenerator.com and Other Sites

The Fake Name Generator web site has an excellent "Advanced Options" page at https://sec487.info/3e that allows you to choose many attributes of the false persona that it will create. Provide the gender, nationality, and age range of the sock puppet you wish them to create and, in seconds, it will return a complete profile to you. The site will randomly generate names, addresses, phone numbers, family members, schools, and even a credit card number for the false persona. According to the FAQ (Frequently Asked Questions) on the site, while the data it returns is randomly generated from word lists, it could accidentally return data for a real person, phone number, or address.

Another, similar web site, https://randomuser.me/, provides some of the data that Fake Name Generator does but is not as configurable. A site on the Tor dark web network, http://elfq2qefxx6dv3vy.onion/fakeid.php, can also assist in generating fake persona information.

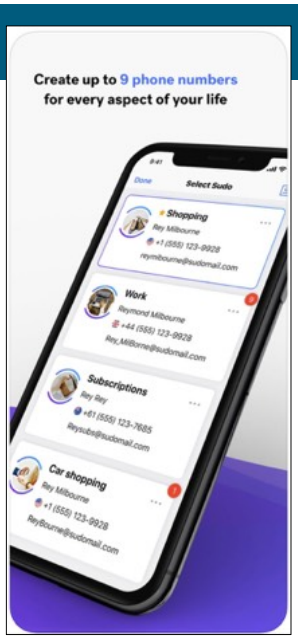
Image from https://sec487.info/jz, January 8, 2017.

MySudo App

- Free/paid
- Up to 9 profiles/avatars
- VOIP calls with phone numbers
- Text for free
- End-to-end encrypted email
- iOS only, for now

ANONYOME LABS

MySudo



MySudo App

There are mobile device applications that can make creating and maintaining sock puppets much easier. The free/paid, Apple-only application, MySudo (https://mysudo.com/) allows its users to create multiple profiles, each with their own sock puppet name, phone number and email address. The phones and emails can be used to register for social media and other web site accounts.

Images from https://mysudo.com/ and https://sec487.info/k0, November 9, 2019.

Overall Rules for Making "Good" Sock Puppets

1. Think about what a **"normal" user** on the platform does and emulate them
2. Use an **unmodified browser** with few plugins and custom settings
3. Try **not to use a VPN** or use a consumer-grade one
4. Use a **cell phone** to make the accounts
5. Use **common services** like Gmail, Hotmail, and Yahoo mail for email systems
6. Practice good tradecraft and only use a **specific network connection method** for a single sock puppet

Overall Rules for Making "Good" Sock Puppets

Creating sock puppets is getting more challenging every day. While the user in me appreciates that social media and other platforms are implementing stronger anti–sock puppet/bot systems, the OSINT analyst does not like this at all. You may have issues creating sock puppets or may create them and then have them disabled later. It is part of the business. There are some practices you can use to increase the chance that your fake accounts will be more accepted by the platforms you make them on. The slide above details several of these axioms. The most basic rule of all is "Look like a normal person and not a security-aware, privacy-focused OSINTer."

Sock Puppet Avatars

Each sock puppet needs a profile image

We can use images from sports teams, TV shows, scenery, or another image that fits your puppet

Some sites may want an image of a face..."your" face

Cannot be a face of a person known to the site

You have options:

1. Find a face on the internet and use it
2. Crop a single face from a picture of a group of people
3. Source your own face images
4. Have a computer "make" a face

Sock Puppet Avatars

One of the main properties of a social media profile is its avatar or profile image. We see a huge variety of things people choose to use for these pictures, including images of their children and pets, logos and players from their favorite sports teams, and picturesque scenes from their travels or destinations they may wish to visit.

But some social media sites will demand an image of your research account's face. This presents us with a problem since our made-up profile does not have a face. We have some choices to source those facial images that we can submit, but many have an important problem. The photo you submit cannot be already known to the social media platform. Many people visit an online image-sharing site and download a random picture of a person to use for their sock puppet profiles. But what if the picture you downloaded was a person's face that already had a social media profile on that site? When you upload the image, your account may look suspicious because someone else is using that image.

We have another alternative: allow a computer system to make a unique image of a person's face—someone who has never existed. Let's explore this option a little bit more on the next slide.

ThisPersonDoesNotExist.com

Using a Generative Adversarial Network (GAN),[1] this website creates unique human faces every time you reload the page

Download, alter, and use these for your profile images

ThisPersonDoesNotExist.com

You need unique human face images? This site has them…well, it actually makes the images. Using a Generative Adversarial Network (GAN), two neural networks of computer systems create and compare generated images to a model (a human face). Images that are close to the model are shown on screen and can be downloaded, altered, and edited to be used as profile images.

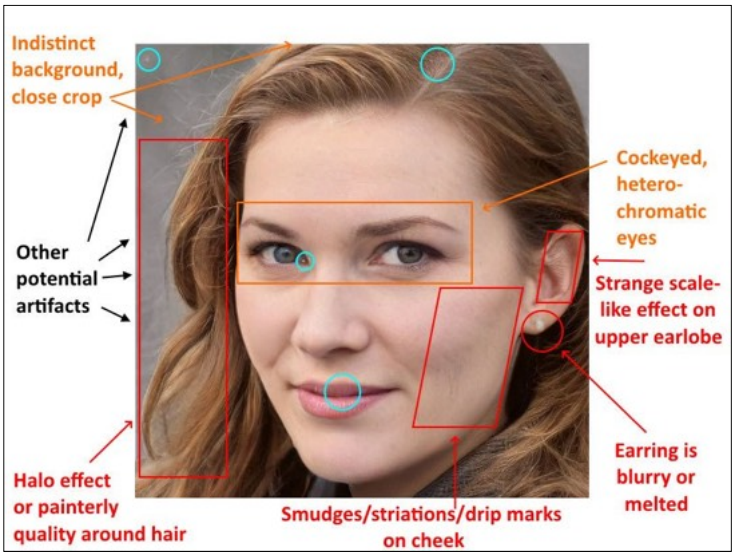
Reference [1] from https://sec487.info/q4, August 9, 2019.

Image from http://thispersondoesnotexist.com, August 9, 2019.

ThisPersonDoesNotExist.com "Tells"

There are pieces of these images that can give away that a computer generated them

June 2019, Raphael Satter (@razhael) described some of them in his Twitter thread[1]



ThisPersonDoesNotExist.com "Tells"

Modifying these images so that they are different from the originals found on the site is important, as there are some artifacts and consistencies that make identifying thispersondoesnotexist.com images simple. Raphael Satter (@razhael) described some of the techniques he used in his Twitter thread.[1]

Some of the "tells" are indistinct background, inconsistent ear shapes, and eyes that are in the exact same position as every other image from this site. Try this: load up the site and put a Post-It® Note on your screen where each eye is. Then refresh the web page several times. Each time a new image appears, I'll bet your Post-It® Note still covers the eyes of the image. Right?

Our lesson from this is, we can use these images but need to alter them by cropping, shading, gray-scaling, and possibly rotating the images so they don't look like every other image from the site.

Reference [1] and image from https://sec487.info/q5, August 9, 2019.

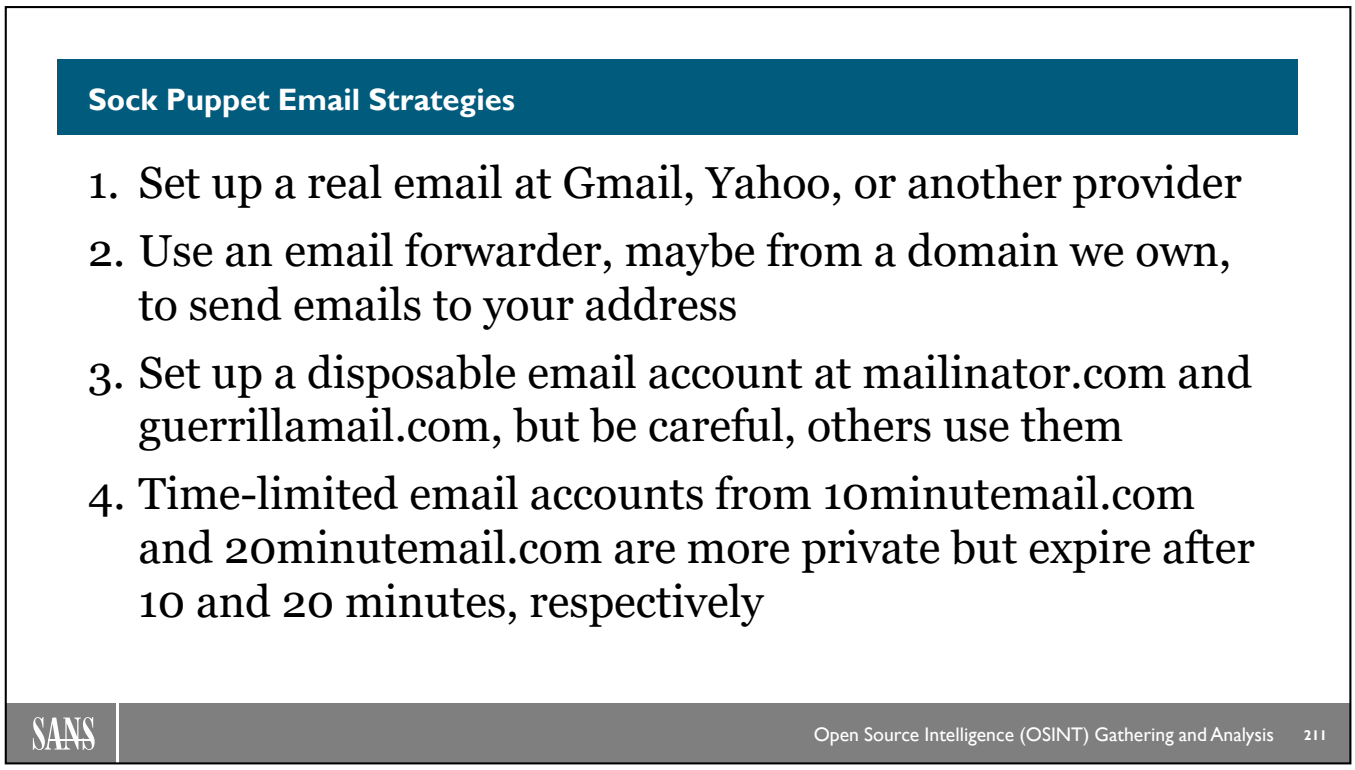
Profile Validation Issues

- Web sites are wise to fake user accounts
- They try to prevent their creation by requiring valid email address verification
- Sign up for an account, and the web site may send you an email with a link or code
- If we created the sock puppets with purely false data, we will not be able to verify it
- There are techniques we can use to bypass verification!

Profile Validation Issues

OSINT investigators are not the only people who make false user accounts on web sites. To reduce the ability of people to successfully gain access to a web site's resources using sock puppets, some web sites will require that users enter valid email addresses when creating profiles. They use these email addresses to send a validation link or code and have the user follow the link or enter the code into the web site. If your sock puppet uses a false email address (for example, dreadpirateroberts@example.com) and the web site sends a validation link to it, you won't receive the code and won't be able to verify the user's information. The web site then can reduce your permissions, deny your sock puppet access, or take other actions until you validate the email address.

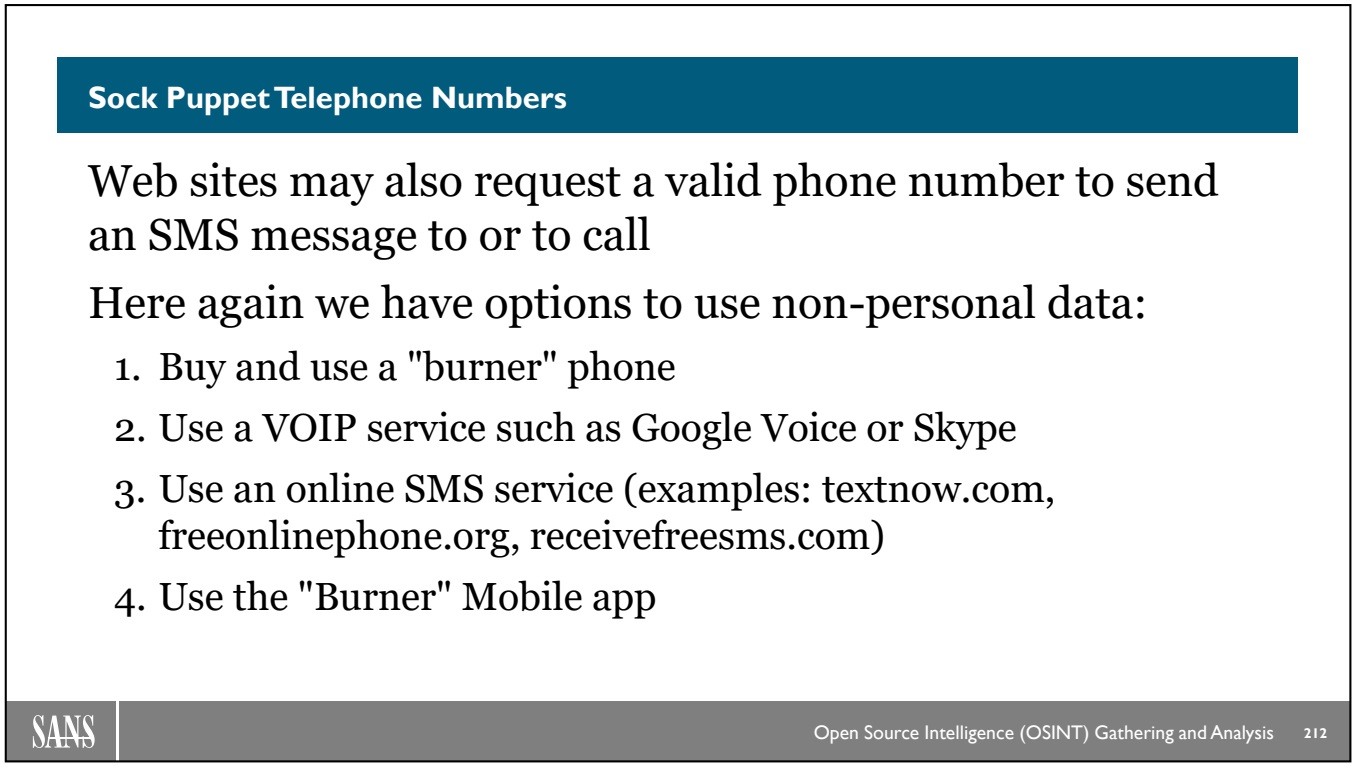
This can be a barrier sometimes, but OSINT investigators are resilient, and we have several options to circumvent these checks.



Sock Puppet Email Strategies

We have several options to get those validation (and other) emails that our sock puppet needs to receive.

1. We can set up a real email account at Gmail, Yahoo, or another provider and then log into it to retrieve the emails. This can work, but some of those email providers may require you to validate your email account by tying a phone number to the account. Now we run into the problem of having to get a real phone number (which we will cover shortly). Some email services do not require further validation to create and use accounts. These would be preferable if you are trying to remain covert.
2. Another option is to use an existing domain or service to set up an email forwarder that receives emails and then forwards them to an account you control. This allows you to keep the real email address you have a secret from the sock puppet's false profile while still being able to retrieve emails. Many domain registrars such as 1and1.com and GoDaddy.com provide email forwarding with any domain you manage through their services. There are also web sites that will set up forwarders from their domain for you. Sites such as securemail.hidemyass.com, meltmail.com, and trashmail.net are examples of these sites. Remember that, if these sites receive your emails and forward them to your real account, they may keep logs of the emails, read the mail, or perhaps tamper with it before it gets to your final email account.
3. If you only need an email account temporarily, then you might choose to use a site such as mailinator.com or guerillamail.com. They have email accounts that you can send validation codes to and then retrieve email WITHOUT AUTHENTICATION via a web interface. And, yes, if you can retrieve email by just visiting a web site, so can others. Additionally, unless you pick a unique email address, you might find that someone else is already using the email address you picked.
4. Lastly, you can get a more-unique email address but one that is only in existence for a certain period of time, such as 10 or 20 minutes. Web sites such as 10minutemail.com and 20minutemail.com will allow you to configure an email address that expires after 10 or 20 minutes, respectively. Others can still retrieve your emails, but this is a little more secure than the temporary email accounts that never expire.



Sock Puppet Telephone Numbers

Some of those web sites you may need to create sock puppets on may require a valid phone number so that they can send validation SMS/text messages or call you to ensure that this is a real account. There are several methods to circumvent this validation step.

1. We can purchase a "disposable" cell phone from a store, the airport, or online and use that phone's number for the sock puppet account. There is a cost (sometimes monthly or yearly charges) involved beyond just buying the phone. Prepaid or Pay-As-You-Go plans can be helpful as they many times do not require a contract.
2. Next, we can use a Voice Over IP (VOIP) service, where we can receive incoming phone calls. Services such as Google Voice and Skype are big names in this area, but there are many other providers that can be used.
3. You might choose to use an online SMS service where you pick a VOIP phone number from a list on a web site, provide that to the sock puppet's web site, and then watch a web page for the incoming SMS message. While this may sound ideal (free, easy to set up, anonymous), there are drawbacks. First, anyone can see the SMS messages that are sent to the service. So, if you choose the phone number 555-555-5555 to receive your SMS messages on, anyone else who is watching the SMS messages for that number will also receive your SMS message. The second problem with this technique is that some web sites know about these services and will not allow you to enter them into validation phone number fields. The social media sites linkedIn.com and vk.com both have used this tactic in the past.
4. There is an application called "Burner" (https://sec487.info/3f) for mobile devices. This app creates virtual phone numbers that you can use on your device.

There are variations and additions to the above list, too. For instance, you could buy a cell phone that allows you to switch the SIM card out. Using a variety of prepaid SIMs, you could change the phone's identity. Another thing you could do is request a voice callback instead of a text message. Then you can use a hotel room, landline, or other number for that verification.

Sock Puppet Long-Term Problems

Reasons for sock puppet issues

- Sites periodically validate user accounts and may discover your sock puppet
- Researchers look for bot-like/sock puppet accounts
- Someone might report your puppet and your account gets banned

Suspicious activities/use

- Lack of site use
- Lack of profile information
- Lack of avatar/profile pic
- Following many but followed by few

Sock Puppet Long-Term Problems

When you create a sock puppet account, an imaginary timer starts, counting down the months, weeks, and days until your sock puppet account is discovered, questioned, and possibly banned from the site. This is a regular battle that we, the sock puppeteers, face.

You can maximize your time with your puppet accounts by better understanding what the triggers are that could get your puppet discovered. Here are some of the major issues:

- Some web sites require reauthentication and reverification periodically. If you sign up for an account with temporary credentials (such as one-time emails) and cannot remember what email you used, you might be discovered.
- Security researchers go "bot hunting" and look for profiles with certain bot-like characteristics. Some of these include:
  - Lack of site use, such as not joining groups, submitting content, and "liking" pages
  - Incomplete user profile information
  - Default avatars and profile pictures
  - Behaviors such as following many accounts but having few accounts (or no accounts) follow back
- Someone might report your sock puppet account to the web site as suspicious for some reason.

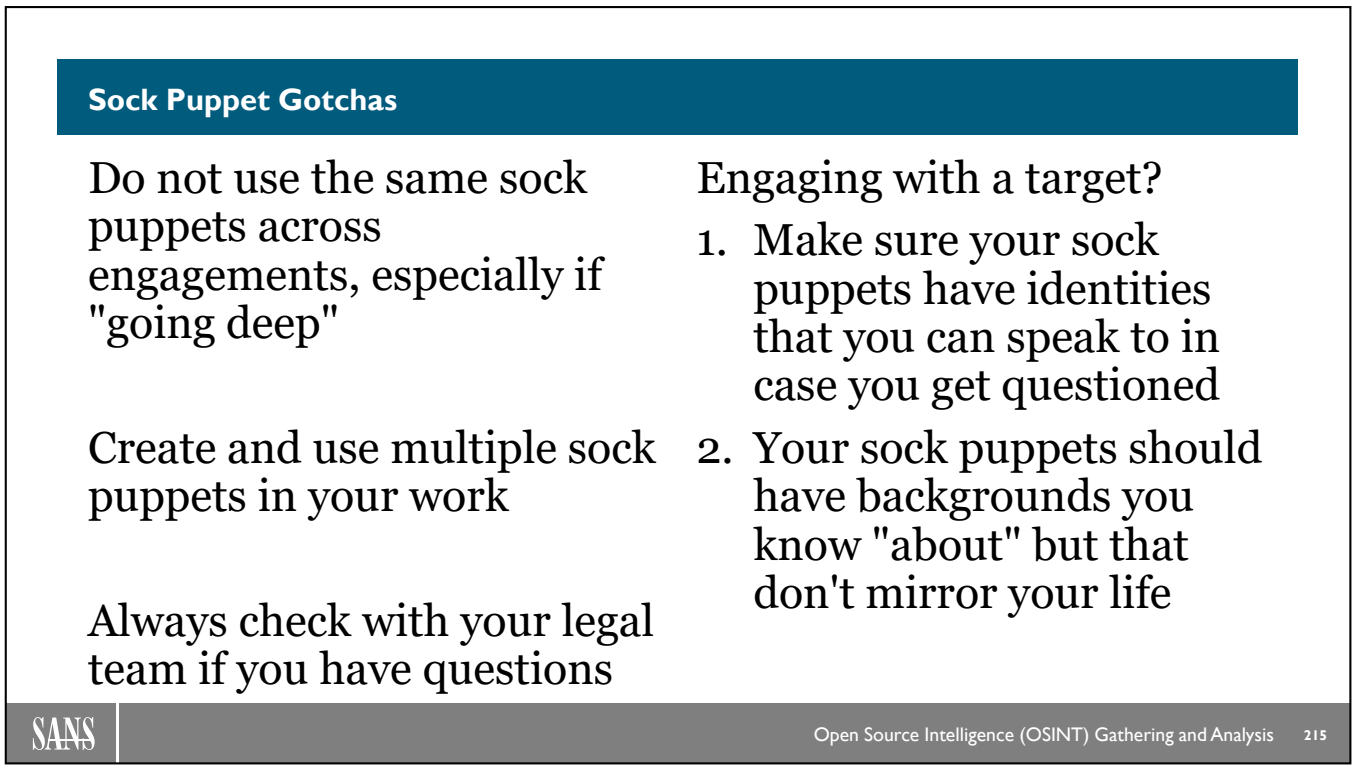
Sock Puppet Long-Term Success Tips

Overall, the more we can make our sock puppets look and act like real people on the sites we create them on, the longer they will go undetected. To do this, you will need to perform regular updates on them. Some tips of actions to perform include:

- If the site has games and third-party apps, turn them on and play away! Playing games has the added benefit of meeting other random accounts on the site. Those people may be willing to be your sock puppet's friends since you have played with them.
- Connect and friend other accounts. If you have an account on a social media site with no friends and connections, it looks new. If you make a bunch of OSINT queries using that account, it makes the account stand out.
- Join groups or teams. Make your sock puppet have depth of character by connecting to their favorite sport's team or TV show.
- If you can, share, retweet, and repost content on the site. A profile with no posts looks suspicious (or abandoned).

You also must be ready to make new sock puppets, as your accounts will get disabled and deleted over time. Create a process to follow to make it easier for you and your team.

Lesley Carhart (@Hack4Pancakes) mapped her journey in the world of sock puppet creation in her blog https://sec487.info/8a. It is an easy read if you would like to avoid making mistakes and getting a bit more advice before making your puppets.



Sock Puppet Gotchas

There are some additional caveats you need to understand before using sock puppets in assessments. This section is more relevant when you are going to use a sock puppet to engage a target long term, but some of the content holds true for the casual sock puppet user, too.

Consider using certain sock puppets for specific investigations. Segment them so that they do not cross over your customers and the work you do. Having multiple sock puppets that represent a variety of "people" to choose from during your OSINT investigation can be helpful for hiding in plain sight. It might not be suspicious if five user accounts tried to look at several gang member's profiles, but a single account doing so might be.

Engaging with your target on a message board, in social media, or in an app can be a valuable tool. If, however, you do not know your sock puppet's history well (or don't HAVE a history for the account) then you can quickly be discovered as a fraud. Ensure that you know answers to questions about your sock puppet's past, where they worked, who they knew, what they saw. There is a scene from the movie Ronin (https://sec487.info/e2) where Robert DeNiro's character asks a person who said they were in the Special Air Service (that is stationed in Hereford) "What's the color of the boat house at Hereford?" (https://sec487.info/e1). Someone that had been stationed there would have been able to answer the question but the character, who had not, could not and his cover was blown.

ALWAYS check with your legal team before an assessment and during it if you have questions. We keep repeating this because it is very important.

Lastly, ensure that, while your sock puppets have stories that you can speak to, they should not have life experiences that mirror your life. They could have some experiences like your earlier life, but it might be simple for a target to perform OSINT on your sock puppet. What would happen if, when they researched your sock puppet's identity, they found your profile that went to the same school as the fake account, grew up in the same town as the fake account, and maybe worked at the same company. If we found this out when doing our own OSINT assessment, we would flag it as suspicious. Many of our targets are smart and would do the same.

SEC487 Day 2 Summary

Today was about collecting and analyzing data.

We explored the harvesting web data, using frameworks, and gathering user names, street addresses, phones, and emails.

Tomorrow we will move into social media.

This page intentionally left blank.