487.1
Foundations of 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.

SEC487.1 | Open Source Intelligence (OSINT) Gathering and Analysis

SANS

Foundations of OSINT

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

TABLE OF CONTENTS

	PAGE
Course Introduction	3
Understanding OSINT	21
Goals of OSINT Collection	32
Diving into the Collecting	62
Taking Excellent Notes	76
Determining Your Threat Profile	103
Setting Up an OSINT Platform	118
Effective Habits and Process	146
Leveraging Search Engines	165

487.1 Table of Contents

This table is a reference for you to quickly move to certain topics in this 487.1 book. You will see a table like this one at the beginning of each book to use to reference the contents of that day's course.

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

FOUNDATIONS OF OSINT

1. **Course Introduction**
2. Understanding OSINT
3. Goals of OSINT Collection
4. Diving into the Collecting
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. Leveraging Search Engines

This page intentionally left blank.

Overview of Daily Content

- **Day 1** - Foundations, goals of OSINT, documentation, sock puppets
- **Day 2** - Harvesting web data, OSINT frameworks, basic data, user names, images and reverse image searching
- **Day 3** - People search engines, social media (Facebook, LinkedIn, Twitter, Instagram) geolocation, metadata
- **Day 4** - Remote location recon, IP/DNS/Wi-Fi, Recon suites, government data, business OSINT
- **Day 5** - Dark Web (Freenet, I2P, Tor) data dump sites, international OSINT, vehicle searching

Overview of Daily Content

We have a very busy week ahead. The slide above outlines some of the overall topics we will cover.

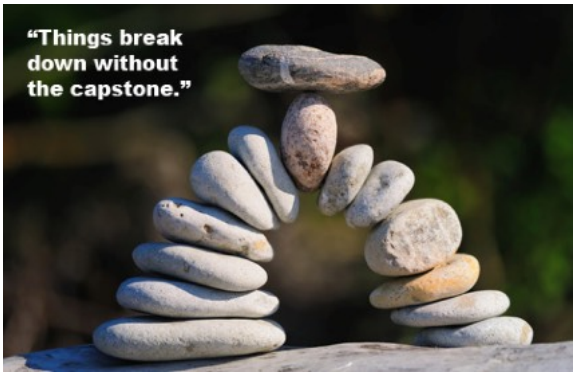
Day 6: The Capstone

You will work in groups to collect, analyze, and present data about a target

The class will vote on the winning team based on criteria

- Breadth and depth of data
- Does it address the requirements?
- How pretty is the presentation?

Fun will be had!



Day 6: The Capstone

After five days of working through process and learning tools, we want you to apply those skills in a real-world event. You and your classmates will divide into several teams and be assigned a target to research. Leveraging the techniques you learned in class (and those you came to class with), your team will perform a complete assessment of the target. Requirements for the collection and analysis effort will be provided by the instructor.

Students will have time to perform their research and analyze the data as a team. Then, each team will decide on how they will orally and visually present this data to their customers (the other students in class and the instructor serve as the customers). Each team will be given points by their "customers" according to a predefined rubric. Higher scores should be given for those assessments going wide and deep in the data discovery. Points may be deducted for assessments not addressing the customer's initial requirements. Finally, more visually appealing presentations may be awarded additional points as well. Each person on each team needs to contribute equally to the effort, or the instructor may deduct points.

At the end of day 6, a winning team will be chosen and prizes will be awarded. Regardless of which team "wins" the event, all teams are required to gain experience in conducting OSINT assessments, apply their knowledge, and have fun. People not having fun will be shown David Hasselhoff's "Hooked on a Feeling" video (https://sec487.info/6) until they smile.

Image from https://sec487.info/je, August 21, 2016.

The Internet as a Storage Medium

- Vast quantities of text, audio, and video are uploaded to the internet every minute [1]
- Much of this data can be personal in nature
- Once something is sent to the internet, it cannot be fully removed

Overview of Topics in SEC487

The internet is no longer a "super highway." It is a place for the digital archiving and storage of massive amounts of data. Each year, web sites share how many minutes of videos are uploaded to video sites every minute. Couple this with the number of tweets, the pictures posted to Instagram, and the number of people who have accounts on Facebook/WhatsApp/VK, and anyone can see that there is a tremendous outflow of data from our electronic devices to a more-public place. The graphic above (available at https://sec487.info/rm) further illustrates the issue. If you look at the infographics from previous years, you can easily see the rapid and large amount of growth, year over year.

In this course, we care less about what news sites are sending to their internet-facing web pages and more about what people are posting. Many times, this data is available to anyone on the internet without restriction. Whether the user has not enabled the security and privacy settings for the site they send their data to or whether the site merely does not adequately protect the user data does not matter to the consumers. We are the consumers. We are the ones who search the internet for bits and pieces of information about a target or targets. We are the ones who know that once something is sent to the internet, it is extremely challenging (if not impossible) to remove it.

Reference and image from https://sec487.info/rm, September 6, 2019.

Privacy? What Privacy?

Privacy settings

- Web site privacy controls may be disabled by default
- Or web sites may have no protection for users
- Many web sites have flaws that bypass privacy settings

Social media sites urge us to “connect” to everyone, creating webs of people

- A person with a less-secure profile can leak your data
- Think about coworkers, friends, and family

Privacy? What Privacy?

When the World Wide Web began, web sites were islands: they didn't share data with, pull content from or authenticate users for other sites. The world was a simpler place with each web site owner storing data locally on their own servers. Contrast that with modern times where we authenticate in one web application, post photos in another and then use a third to "mash" them all together into a single page for our connected friends to see.

What protects all of this data? Each site you create an account with or send data to may have its own privacy policy and security controls to limit who and what can access your content. But this varies from system to system, and controls can even change within a site so that what you protect now may become unprotected when the next change to the application is made. Some web sites have no method for users to control who has access to view data about themselves, and other web applications may have flaws that allow unauthorized people to collect users' info.

Compounding these issues, social media sites ask their users to connect, friend, and link to other people. These relationships to children, spouses, family members, work colleagues, and others are important to OSINT analysts. Sometimes we can retrieve information about a target by finding a connection to them that has not enabled the privacy and security settings. If the target posts something to their circle of friends but one of those friends allows people outside their circles to see their friends' posts, the target's post may be viewable by anyone (including us)!

Data Out of Our Control

Even if you don't use social media, others do

- You appear in a photo, and facial recognition matches you

There are other sites where users have no ability to prevent data from being published

- Government data – judicial issues, registrations
- Real estate info – taxes, who bought and when
- Religious groups and hobby club newsletters
- Life cycle events – births, deaths, marriages, divorces

Data Out of Our Control

If you try to be a "digital recluse" and not participate on social media sites and not use email or cloud services, you may only be decreasing your digital footprint a little bit. There is data about us that makes its way to the internet even if we don't send it. Facebook and other sites use facial recognition to match people to their profiles. Take a picture with me in it and, if I am your "friend" on Facebook, the web application may suggest that I was in the picture. Take this information and think about all the pictures that you took on your last vacation. I am sure that there were other people in those pictures; probably people you don't know. They are just "extras" to us, but if they are on Facebook, Facebook (and others that have access to Facebook's backend systems) may know who they are.

Much of the business that happens in government can be viewed and searched on the internet. Court records, real estate listings, and business filings are searchable and provide reliable data to those searching for it. While I can alter the information on my LinkedIn page, I do not have access to change that parking ticket I received last year. Because we know our targets cannot change government data and it is reliable, it is highly sought after during investigations.

Belong to a religious organization that publishes a monthly newsletter in a PDF that is posted to an unprotected web page? Your information may be published without your knowledge when you have a birthday, a wedding, or even a death in the family. These sites are excellent third-party sources for OSINT.

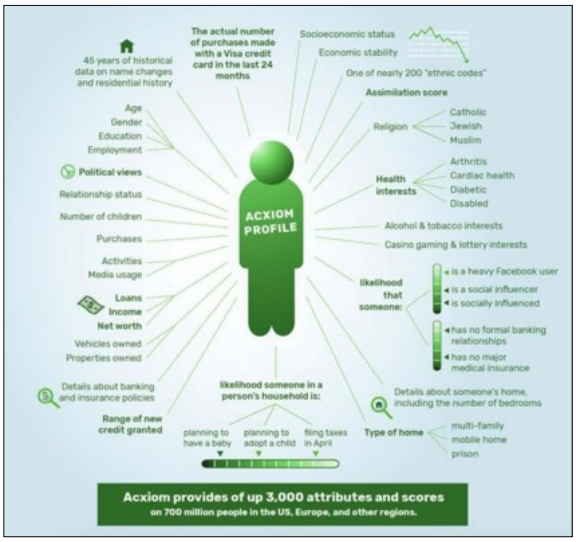
Do you run races? Running, biking, and other race results are excellent places to get full names, ages, genders, and possibly family and friend connections from (because the people who cross the finish line with you may be your friends and family).

Data Brokers

Within the United States, companies can buy and sell data about people without first asking permission

The image on the right shows some of the types of data companies collect

This data gets sold and aggregated



Data Brokers

There are companies within the United States that know a lot about the people who live and travel within that country. These data warehouses or data brokers buy, collect, analyze, and sell huge amounts of data about people and their activities to financial firms, advertisers and marketers, and people search engines. Using many different techniques—from cookies that are tracked within web browsers, to applications installed in smartphones, to public information—these companies assemble profiles and dossiers on millions of people, including their likes and dislikes, their health issues, and their activities.

The profiles they create can be purchased by other companies or people who use them for everything from selling products to influencing elections.

Reference and image from https://sec487.info/q1, July 21, 2019.

Connecting the Dots

- A large amount of data posted about people
  - Public data
  - Posted with weak or no security
- Data is online 24 x 7 x 365 (all the time!)
- Information out of our control
- This means significant opportunity for collection, analysis, and use of OSINT
- "Target rich environment"[1]

Connecting the Dots

Putting this all together, we have a large amount of information that is either poorly protected or unprotected. It is retrievable by anyone around the world, 24 hours a day, 7 days a week, 365 days a year. There is data that is sent to the internet about us that we cannot prevent from being posted. Add to all of this the voluntary posts, photos, and videos that people who use exercise, dating, networking, and social media sites send, and you have a "target rich environment."[1] These are the easy parts of the OSINT analyst's work: collecting and analyzing data from the surface web of the internet.

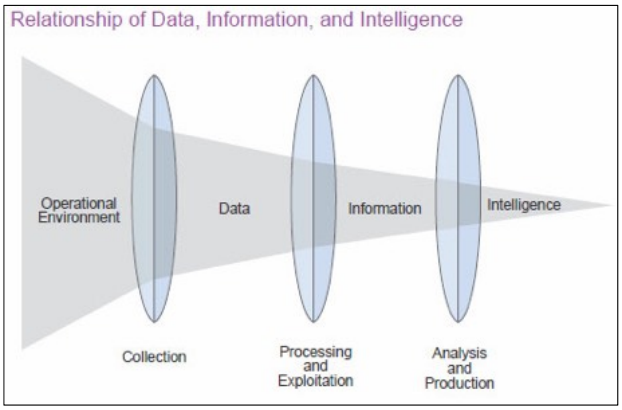
Reference:

[1] A "target rich environment" is a quote from the 1986 movie *Top Gun* (https://sec487.info/5).

Data, Information, and Intelligence

The United States Joint Intelligence Publication and the NATO OSINT Handbook v1.2 describe:

1. Raw data
2. Processed information
3. Analyzed intelligence



Data, Information, and Intelligence

The United States Joint Intelligence Joint Publication 2-0[1] from 22 October 2013 and the NATO OSINT Handbook v1.2[2] use similar terms to describe the differences between open source data, information, and intelligence. Since our goal in this course is to present our customers with intelligence, let's break down their graphic into its three main sections:

1. Open Source Data (OSD) is raw content that is gathered or collected. With the understanding of where it came from (its source) and with other data to provide context and meaning, we can perform processing tasks on it to transform the individual bits into something more meaningful—information.
2. Open Source Information (OSIF) is processed data that leverages filtering and validation processes. The above diagram explains that it is also "exploited," which means that places of interest within the data are identified that the consumer or customer may find useful.
3. Scrutinizing the information, adding analysis, making recommendations, and publishing the results change information into Open Source Intelligence (OSINT). This is our goal.

References:

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

[2] https://sec487.info/d7

About the Course

Focus on the Process

- This is the framework that helps support our work

Leverage the tools

- These help us move faster and go deeper into the data

Practice all of the above

- Practicing increases our experiences

About the Course

This course was designed to focus on three overall concepts: process, tools, and practice.

As we will discuss in the coming modules, OSINT investigations can start with a simple question, "What can you find out about a target?", and end up leading to courtroom testimony, legal actions, or worse. We are going to work on understanding the process of OSINT collection, what drives it, what the challenges we may need to overcome are, and what questions we need to ask. You will begin to create the framework that governs your assessments: which sites will you visit? Which tactics will you employ? Deciding this in advance of a request for OSINT work will make executing on your assignments faster, more repeatable, and more in-depth.

Tools will assist us in our collection, analysis, and reporting tasks. Some of them are focused on collecting data faster than we can with a web browser. Others aim to collect data from a wide array of web applications. We will also discuss which ones can help us in documenting our work. One thing to note is that not all "tools" we use in OSINT are Python or Ruby scripts that run from our collection computers. Much of our work is spent using special web applications or custom queries to search on web sites. Using the word "tool" to refer to anything that helps us further our investigation will serve us well in our work.

Finally, knowledge without practice is less likely to remain in your memory long term, so we will use exercises after key modules to cement the process and practice with the tools that you will use when you perform your OSINT work. Practicing increases our experiences which we then use to make decisions faster, analyze data more accurately, and ultimately, help our clients more effectively.

Course Learning Level

- This is a foundational course
- It is meant to be a starting point
- We cover tools, techniques, process, and workflow
- Some of you may have years of OSINT experience
- Understand that this course is meant to expose everyone to a wide variety of topics at varying depths
- Some places you might know more and some places less

Course Learning Level

As this is the first dedicated OSINT course within SANS, it is the beginning. It is meant as that first course anyone can take to gain an understanding of the landscape of the OSINT field. It covers a wide variety of topics and does so at varying depths. Some modules of this class will be limited to what can be retrieved from sources using our web browsers. Other modules will take students into a terminal window and have them using the command line to perform OSINT work.

We bring this up so that you can understand the intended audience for the class. For those students who have been honing their OSINT skills for years, this beginning course may only fill in gaps in your knowledge. For those students who are specialists in one subject matter (e.g., the dark web, Arabic social media, etc.), this course will provide you opportunities to expand your understanding of the OSINT world.

Throughout the course, the author and your instructor will seek to challenge you to try new techniques and tools to achieve your OSINT goals.

SEC487 Alumni Slack Group

Slack is a communications platform between instant messaging, a Google/Yahoo group, and a forum

Web, mobile, Windows, Mac, Linux clients (it is on your VM!)

Join: **https://sec487.info/joinslack**
Slack: **https://sec487alum.slack.com**

SEC487 Alumni Slack Group

Sharing data and collaborating across a group of people can be challenging. There are many online platforms that can be used and different formats for the sharing. We have instant messenger apps and forums, Google groups, and Yahoo communities. We decided to use the Slack.com platform to create our OSINT community.

Slack has clients for mobile devices (iOS and Android), Windows, macOS, and Linux. Don't feel like installing a client? That's not a problem. Use the web-based interface to communicate with your colleagues.

We know that both during and after class, you might have ideas to "bounce off someone" or questions to raise to OSINT-minded people. So, we created and now maintain a Slack Group, https://sec487alum.slack.com, for SEC487 students to share ideas and keep the conversations from class going. Join other students and instructors in OSINT-related discussions in this group.

This group is an open group. To join, you (and your colleagues) can visit https://sec487.info/joinslack and sign up.

Images from https://slack.com/features, December 9, 2017.

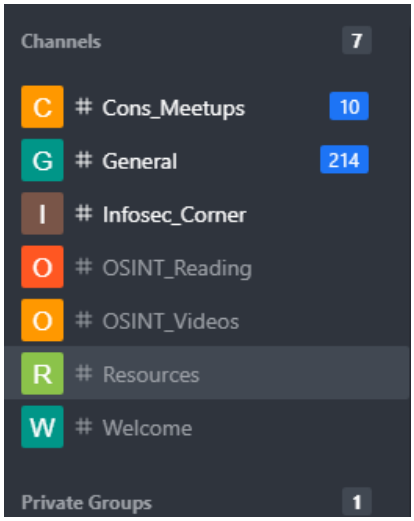
Open OSINT Rocket Chat

Large community of people of all skill levels

Persistent chat rooms, CTFs, questions, resources, and more

Some deeper, private channels too!

International scope and user base



Open OSINT Rocket Chat

The Open OSINT Rocket Chat group is a supportive and knowledgeable place to learn OSINT and help others. Big names in the industry participate in the online chat discussions. This free resource is open to anyone and is not associated with SANS or this course. Sign up for a free account at https://osint.team and use the Rocket Chat application or another to access the chat.

There are channels that all users have access to, such as privacy and resources, and then there are deeper, private channels where people discuss dark web issues and work on real-time OSINT issues as they develop.

People from around the world are in this group, and that provides a variety of inputs, ideologies, and motivations.

Long URLs and Short Ones

- We use a short URL generator with the **sec487.info** domain
- This is done to save you time typing
- We will not give you a knowingly malicious URL

Link: https://www.imdb.com/title/tt0093779/
Shortened[1]: https://sec487.info/princess

Long URLs and Short Ones

This OSINT course harvests data from the internet. As such, there will be many URLs that we see in slides and in the notes pages. We provide all URLs as shortened URLs so that it's easier for you to enter them in your browsers. These are meant to help you work more efficiently. Instead of trying to type https://www.imdb.com/title/tt0093779/ into your web browser, you can use the easier URL https://sec487.info/princess, which will take your browser to the bit.ly web site and then redirect it to the IMDB page for the movie *The Princess Bride*.

Since you are in this class, you are most likely aware of the attackers and scammers that use short URLs to hide redirections to malicious sites. You enter a malicious short URL into your browser, it gets redirected to the attacker's site, and then your web browser gets attacked and possibly compromised. This is not something that the author or SANS want for you. I'm going to write "You can trust the short URLs that are in this course," now knowing that some of you will not want to use them. That is totally fine. Choose your own method of visiting the URLs.

[1] Yes, we understand that the domain sec487.info is not really "short". However, it is easy to remember and should make your job getting to different web resources easier.

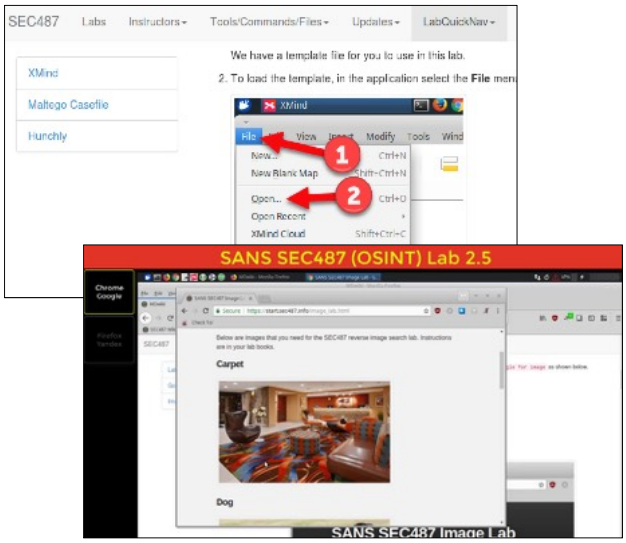
Digital Workbook and Videos

Digital workbook inside VM contains:

- Digital versions of labs
- Videos of lab walkthroughs
- Tool and technique descriptions

Update workbook content via a terminal window and:

```
$ sudo wikiup.sh
```



Digital Workbook and Videos

To make a more interactive experience for students, we include a digital workbook inside your virtual machine. It has helpful tools and technique tips and complete electronic versions of all course labs. The workbook can be especially helpful to students when performing the labs, as they can cut and paste from the electronic documentation (or click hyperlinks) instead of having to type long strings of content into browsers or terminal windows. This allows students to focus on the OSINT techniques instead of their typing skills.

Also included for each lab inside the workbook are video walkthroughs. We understand that you have a lot of learning going on while you are taking the course and, sometimes, it is helpful to see someone else go through the lab. The videos not only do this but sometimes contain additional interesting comments and tips.

The workbook and its content can be updated from a terminal inside your VM. Simply type sudo wikiup.sh and then enter your student user password, and the script will update your system. Note that your system needs internet access to pull content from GitHub.com for this to work.

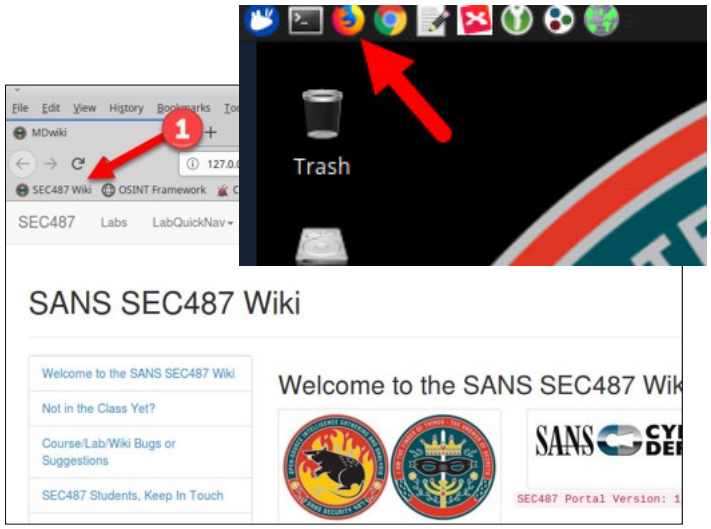
Using the Class Virtual Machine

Username: `student`

Password: `Security487`

Workbook:

- Launch Firefox
- Workbook is home page or the first bookmark in the bookmark bar (arrow 1 at right)



Using the Class Virtual Machine

In this course, we use a virtual machine (VM) for our labs and as our working platform. To log into the VM, use the username and password from the slide above.

To reach the electronic workbook for your class exercises, launch the Firefox web browser (see image above). The workbook contents are the home page for the browser or can be accessed manually by clicking the first bookmark in the bookmark bar (arrow 1 in the image above).

WARNING! "Here be dragons"[1]

- The internet has content that is upsetting to some people
  - As an OSINT analyst, you **will** come across this content
  - In this class, you may encounter it in your exercises
- The author and SANS have no control over the content on the internet and what may appear in your labs
- Where possible, we have made labs "safer"
- We are adults and can handle this
- We will try to minimize any unpleasantness

WARNING! "Here be dragons"[1]

The internet is not the simple, safe place it was in the 1990s, when static HTML content, hit counters, and dancing Homer GIFs were common. It is common for web sites to host malware, pornography, and graphic, violent images. As OSINT analysts, we cannot know where our searches will take our browsers during our assessments. It may be that to answer your customer's questions, you may need to visit some of these darker places on the internet.

SANS and the author of this course do not wish to upset anyone and, yet, we realize that it is important that we teach real-world skills to our students. Doing so means that the exercises you will be doing will be on the live internet. Because of this, you may encounter photos, videos, and/or web content that may be appalling or upsetting. If you do, please close the browser page and move on in your investigation. Choose a different site for the labs in this course. In your professional OSINT investigations, you will need to determine what you will do when you come across these sites. Build it into your processes and inform your clients of the boundaries of your investigations.

Reference:

[1] "Here be dragons." Wikipedia. https://sec487.info/dragons, July 27, 2016.

Current URLs

- Sometimes this courseware cites URLs that were valid at the time originally cited
- URLs change over time
- This courseware does not necessarily update all URLs
- Out-of-date URLs still have value to us.
  - Hints about how to find source material
  - Can be used to find material at the Wayback Machine maintained by the Internet Archive

Current URLs

We make every effort to ensure that the URLs and content of this course are as up to date as possible. However, you may notice that there are some URLs in the courseware that do not lead to web resources. We have left these remnants of the past in the courseware, even though the resources may no longer be available, to help you go look for the resources. Sometimes we can use an old path to a file or web page and retrieve the original from a web caching source, such as https://archive.org.

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. **Understanding OSINT**
3. Goals of OSINT Collection
4. Diving into the Collecting
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. Leveraging Search Engines

This page intentionally left blank.

What Is OSINT?

OSINT is Open Source INTelligence

Information in the public domain or accessible from public sources

- Media such as audio, video, and pictures
- Text from documents, articles, and blogs
- Maps and geolocation of data

Social Media

- Called SOCMINT for SOCial Media INTelligence

What Is OSINT?

Open Source Intelligence, or OSINT, is the process of searching for, gathering, and analyzing data found from public sources. Most of the time, this data is in the public domain, but sometimes information behind a paywall or requiring some form of authentication can be helpful as well. While we focus on OSINT that is easily accessible on the internet, you might find that visiting a courthouse and examining records and using printed resources at a library can help your OSINT assessments move forward, too.

During the course of a normal day, many of us use common web sites that an OSINT analyst would visit to harvest data. Photos of a certain location or person, a video blog (vlog) entry that was recorded inside a target's office, and a post that a target made in a forum are all examples of data that OSINT analysts collect.

When we are working with data about people, some of the most important content to gather and examine are what the target is posting, or what someone else is posting about them, on social media sites such as VK.com, Facebook, and Twitter. Social Media Intelligence, or SOCMINT, can, depending upon the target, yield an overwhelming amount of data to harvest and analyze.

Data Is Just Data until Analyzed

- Anyone can google a topic
  - As OSINT people, we regularly look on page 2 or 3 of a Google search result page
- The collection, organization, and analysis of that data is important
- The "intelligence" portion of OSINT is the key
- And there may be a LOT of data to sort through

Data Is Just Data until Analyzed

Some people remark that OSINT is "people just googling things." This short-sighted and inaccurate definition removes several of the most important pieces of OSINT: the organization of the information gathered and the analysis of the data. While it is true that most people can use a search engine to find the answer to simple questions, the OSINT analyst is often asked very challenging questions where there may be no 100% correct response.

When you have seen friends and family members search for something on the internet, how many times do they view the items on the second or third page of results? Most of the time I see people find "good enough" answers in the first 10 response entries. It is rare that people will go to the rest of the search results, and this is why there is such a huge business around Search Engine Optimization (SEO). Getting your product or your company's web page to appear as the first 5 or 10 results can make a large difference in ad revenue and traffic to your web site. As OSINT investigators, we care about entries that can be on any page of the results. That is why successful OSINT analysts focus on collecting everything and then analyzing it for meaning and context.

The intelligence portion of OSINT means that we, the analysts, are an important piece to the process of an investigation. "Just googling it" might get an answer for a customer, but is it a complete answer? Is it the best answer? Later in the course we will discuss effective strategies to use when interpreting data. Depending upon what your customer has asked you to search for, there may be an overwhelming amount of data to collect, parse, and examine.

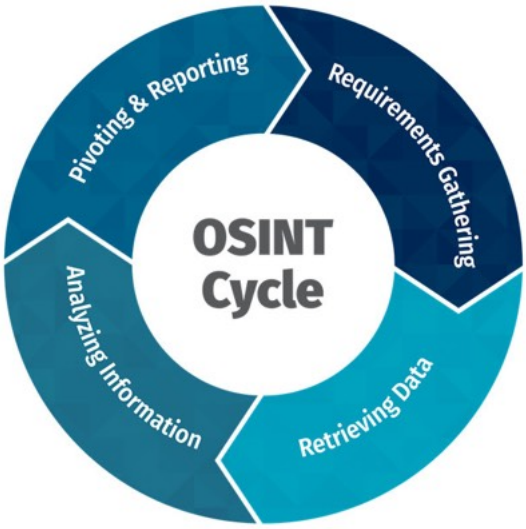
The OSINT Cycle

Requirements gathering

Retrieving data

Analyzing information

Pivoting to a new perspective or Reporting analysis



The OSINT Cycle

While there are many variations on this cycle, most agencies and organizations agree that there are at least four main stages of an OSINT cycle:

1. Requirements gathering begins the process. Here the OSINT analyst learns from their customer what they would like, as well as when and how. This stage is very frequently skipped, as people assume they know what their customer wants instead of asking them for concrete requirements.
2. Retrieving information is the step where an analyst searches for key terms, collects photos and other media, and scours web and database content. This process can sometimes be challenging because the analyst needs to judge and understand when they have obtained "enough" data to analyze.
3. What separates OSINT investigators from people that "just google it" is their analysis. This is the stage in the cycle where the analyst examines the pieces of data they collected, applies their understanding of their customer's requirements, and sorts the data into useful and less useful categories. There will also be other categories such as "follow-up" and "pivot" that will tell the analyst to pursue the topic further in additional queries (also known as going deeper) and to switch their searching to pursue a different topic, respectively.
4. After the OSINT analyst collects enough detailed data on a specific topic (for example, their target's educational background), there will come a time to pivot to pursue a different line of exploration (for example, their work history). Alternatively, if “enough” information has been analyzed, the analyst may move into a reporting phase. After the OSINT analyst collects enough detailed data on a specific topic (for example, their target's educational background), there will come a time to pivot to pursue a different line of exploration (for example, their work history). Alternatively, if “enough” information has been analyzed, the analyst may move into a reporting phase. Now the analyst has additional details and pivot points that they can use to refine their searching. Requesting and/or confirming requirements from their customer is helpful to ensure that time, money, and investigation constraints are being met. The cycle repeats.

Some resources call these stages by other names. For example, https://sec487.info/3i uses harvesting, enriching, and reporting to describe the second, third and fourth stages of the cycle, respectively.

Regardless of the names you give them, there are stages that your investigation will travel through, over and over, during the course of your assessment.

Other Resources on OSINT Processes

(Links to these documents and sites are in the notes)

- Chapter for Strategic Intelligence on OSINT
- Operations Security Intelligence Threat Handbook
- Handbook of Electronic Security and Digital Forensics
- ITACG Intelligence Guide for First Responders

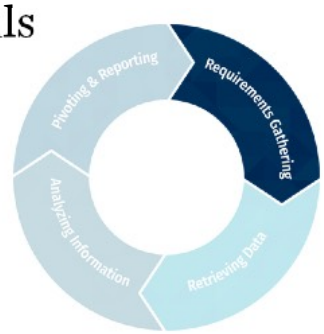
Other Resources on OSINT Processes

Your OSINT process will vary, depending upon where you work and who your customers are. If you are an intelligence officer, you may have a rigid, formal process with inputs and outputs described by military documents. Private investigators may have a simpler process to follow. In any case, there are additional resources you can use if you would like to read about how the OSINT process evolves in different industries. Below are several notable pages and documents:

- "Chapter for Strategic Intelligence on OSINT," https://sec487.info/3l (February 4, 2017)
- "Operations Security Intelligence Threat Handbook," https://sec487.info/3m (February 4, 2017)
- "Handbook of Electronic Security and Digital Forensics," https://sec487.info/3n (February 4, 2017)
- "ITACG Intelligence Guide for First Responders," https://sec487.info/il (August 21, 2018)

Requirements Gathering

Working with your client to discover their goals



Example topics covered:

- What techniques are in scope and what are out?
- Time allotted for OSINT work
- What output do they want?
- How covert would they like you to be?
- Do they have some information that can help you?

Requirements Gathering

The first stage in an OSINT engagement is to meet with your customer and discover their goals for the assessment. This step is often skipped but should not be. Create a document of standard questions that you need answered so that your engagement stays on track and focused on your client's needs.

Some topics that may be covered in this phase are:

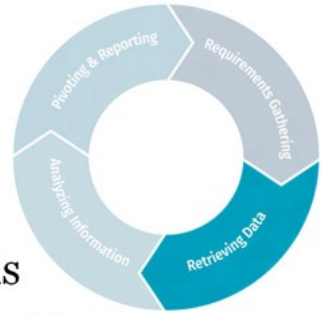
- What techniques are in or out of scope for the assessment? Are you allowed to connect with the target or not?
- How much time do you have to do your work?
- What content and in what format does your customer want the output from your investigation? Do they want a simple word processing document or PDF, or are they looking for a verbal briefing as well?
- How covert or overt would they like your system and network traffic to be?
- Does your customer have information that will help you begin your assessment? Name, address, and phone number of the target? Email address? Usernames on web sites? Dates and time frames they are interested in?

Our customers may also expect that we can hack into other's systems and retrieve private data like they may have seen in movies or on TV. We may need to explain what we can and cannot do for them in this step of the assessment.

Retrieving Data

A straightforward step in the process

Search and collect data about your target



Record and download everything for analysis

- Photos, videos, URLs, text documents, relationships

How will you record what you collected?

Retrieving Data

Once we know what our engagement parameters are, what techniques we can use, and what our client's goals are, we move into the next stage of the OSINT cycle: retrieving data. This is what people think about when they hear about OSINT: analysts using search engines, browsing web sites, and reviewing online documents.

We record everything we do. We collect and download all the data we can discover for later analysis. This will include everything from text to URLs to videos to photos to documents. Since the internet, and the data that is hosted on it, is not something that we control, we grab copies of anything that we want to use in our analysis and our report. Having local copies of data ensures that if a resource is altered or removed from the internet, we still have a copy of it. Sometimes it is these removed pieces of data that are most revealing.

We will cover the tools that we can use to record "all of the things" later in this course.

Analyzing Data

Evaluation of collected data

Transform data into information through analysis

Key points when evaluating data:

- Is it relevant?
- Is it accurate?
- Is it objective?
- Is it credible?
- Is it corroborated?
- Is it a pivot point?

Analyzing Data

Once we have gathered "enough" data ("enough" is a subjective measure), we will move to transform data into information by evaluating and analyzing it. Some key points to use in our review of the data may include:

- **Is it relevant?** Sometimes we gather content that, upon analysis, is not helpful to the overall investigation. It may be interesting but might not be something that matters.
- **Is it accurate?** The internet contains false data. Whether this content is meant to be posted or just something that was generated and is accidentally erroneous, we need to evaluate the bits we retrieved to see where that data falls. We can report information that we suspect may be inaccurate, but we need to ensure it is noted properly.
- **Is it objective?** People posting on the internet sometimes seem to have agendas. They may want to prove something or provide one side of a story. When examining data, we need to understand if the content is fact or subjective.
- **Is it credible?** Anyone can post anything to the internet, so you need to evaluate the data about your target for its credibility. Here is where the place the data came from will be important, as some places generate more credible data than others.
- **Is it corroborated?** Can we find the same data across multiple, unrelated web sites? Just because data is corroborated may not make it truthful. Many web sites draw from the same content. If that data is incorrect, each site will repeat the data and it will not be true.
- **Is it a pivot point?** Analysis will reveal that we may need to perform more research on certain data points. Look for those areas and note them.

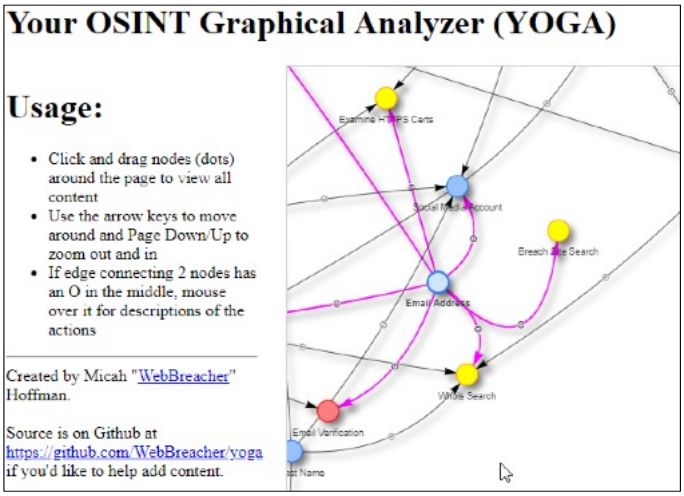
YOGA Helps with Pivots

When gathering data, it is helpful to understand what you can transform it into

Understanding these pivot points comes with experience

Your OSINT Graphic Analyzer (YOGA) helps by showing connections

https://yoga.osint.ninja



YOGA Helps with Pivots

As our OSINT experience grows, we naturally see how to transform data from one type to another. We understand that we can perform DNS lookups on domain names to get their IP addresses (we cover this later in the class). Searching social media sites for a username becomes natural once we figure out those connections. To help show these connections between data types and actions you can perform, Micah Hoffman created Your OSINT Graphical Analyzer (YOGA) and hosts it on GitHub at https://yoga.osint.ninja.

YOGA's interactive display connects classes of data (blue dots) with actions to perform on them (yellow and red dots). You can move the nodes and edges (lines connecting them) around, hover over nodes and edges to get more information, and zoom in and out, all within this JavaScript-based site. Best of all, you can help contribute to the project by suggesting your own ideas at https://sec487.info/hy.

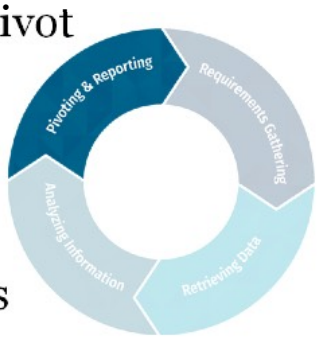
Pivoting to a New Perspective and Reporting

If we need more data about something, we pivot

If we have enough information, we report

Reports have data, information, and analysis

Show your client evidence and tell them why it matters



Pivoting to a New Perspective and Reporting

If the analysis phase of our assessment yielded areas that we need more data on, then we will pivot and start performing our searches on those data points. We will jump to the "retrieval" portion of our OSINT cycle, collect, record, and then move back into the analysis phase.

Alternatively, if we have gathered and analyzed information that satisfies our client's goals for the assessment, then we can move into the reporting stage. Here we will craft a document that contains the initial guidelines and requirements, show the relevant data that was discovered, add our analysis to explain why it matters and what it means, and then describe next steps and conclusions.

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. **Goals of OSINT Collection**
4. Diving into the Collecting
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. Leveraging Search Engines

This page intentionally left blank.

OSINT for "Good"

OSINT data enhances investigations

Let's examine a number of the different groups that use OSINT for "good" or "lawful" purposes

OSINT for "Good"

OSINT techniques are leveraged by many groups around the world to find specific information and to enhance existing bodies of knowledge. An example of this might be a journalist who examines location-based protests through the Twitter feeds of citizens in that country. Or perhaps a business is considering acquiring another company and would like more information about how the public views the company before purchasing it. Through the collection, analysis, and aggregation of public data, groups from law enforcement to parents to businesses can understand situations more completely. Let's explore these categories in more depth in the coming slides.

Image from https://www.pexels.com/photo/flat-view-photography-of-four-persons-sitting-facing-laptop-on-desk-1451447/, August 8, 2019. (URL no longer active)

Law Enforcement Agencies Use OSINT – Criminal Networks

- Why are these associations important?
- Tracking gang and "club" activity
  - OSINT on tattoos
  - Social media profiles, pictures, and comments
- Mapping associations using tools to visualize locations and connections

Law Enforcement Agencies Use OSINT – Criminal Networks

Within the law enforcement world, understanding what people are associated with what groups can be an important link in solving and preventing crime. OSINT helps the police track these relationships through online posts, connections, friendships, pictures, and videos. Some gangs, clubs, and hate groups maintain a high social media presence to publicize their exploits, convey their messages, and recruit new members. Analysts use this information to compile links to other groups and people. Show a picture of four people posing showing gang signs or all with the same club tattoo, and there may be a relationship.

Analyzing the content and context of media posted on these sites can show areas where the gang considers it their territory, who are the current leaders of a group, and other information that law enforcement can record and leverage in their work. It is sometimes useful to connect individuals to others in a network-like pattern to show more intimate relationships. Doing so can connect small cells of people to larger ones and highlight members who may be key to both groups.

Image from "Project No Gangs," Project No Gangs, https://sec487.info/jd, August 27, 2016. (URL no longer active)

Law Enforcement Agencies Use OSINT – Posting about Suspects

- Social media postings self-incriminate
- Show locations and details
- Suspects “Bragging”
- Social media "Wanted Posters"

Law Enforcement Agencies Use OSINT – Posting about Suspects

Law enforcement can also harvest data from open and social media sources. Some of this data may incriminate suspects, taunt them, or just ask the public for assistance.

Image and reference: https://sec487.info/7.

Social Media Self-Incrimination

Some people don't think that the police are monitoring their social media...

...or that their friends won't turn them in to law enforcement

Take Mr. Michael Brown's bragging as an example

Social Media Self-Incrimination

There may be a perception by some people that social media posts are private. Other people just like to brag about their "accomplishments" and, while it may be fine to do so in a private setting with some good friends, doing so on the internet may lead to jail time. Take, for example, Mr. Michael Brown. His story unfolds on Detroit's WXYZ web site found https://sec487.info/8.

Image from https://sec487.info/8, August 24, 2016.

Social Media Bragging

The post, from Brown, reads in part: **"I showed up some hardleys at the bar in orion they made noise I took it a step beyond cop seen me put his car right at my foot. Flicked his lights as soon as I heard his car hit park I dropped into gear and disappeared 45 sec. Chase. 140 in a 35mph."**

He then signed his post with three hashtags, #ftp, #nojailthisweekend #everyonelovedit.

Social Media Bragging

Mr. Brown caused his motorcycle to make some very loud noises in front of a restaurant he was leaving. Police started following him with lights on and, as Mr. Brown self-reports in a Facebook post, "I dropped into gear and disappeared." Fleeing from the police is a crime in most areas, and this would be bad by itself. But Mr. Brown was on a roll and decided to add that his motorcycle hit "140 in a 35mph" speed zone. Freely admitting to his Facebook friends that he went four times the posted speed limit was something that ultimately helped seal his fate.

Upon learning that the police had issued a warrant for his arrest, Mr. Brown turned himself in to the police. It is important to note that most social media posts can be used in court. Ending his Facebook post with "#ftp" (f___ the police) and "#nojailthisweekend" may tell the court system a little about what he thought about his actions.

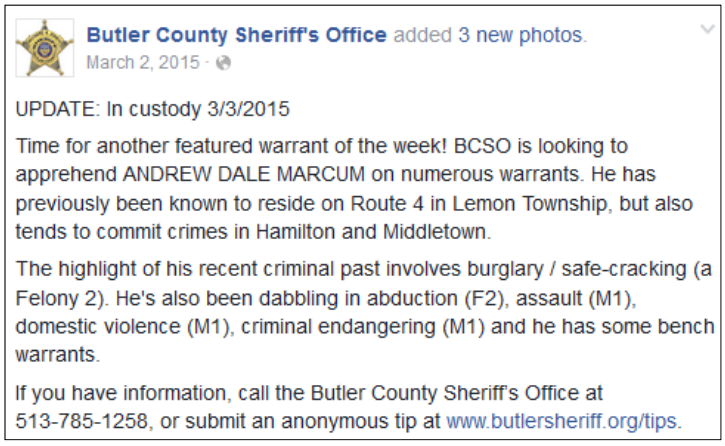
Image from https://sec487.info/8, August 24, 2016.

Social Media "Wanted" Posters

Law enforcement use social media as modern day "WANTED" posters

Ohio's Butler County Sheriff's Office is a great example

They use Facebook to reach a local audience



Social Media "Wanted" Posters

In the United States' recent past, law enforcement tried to spread the word about criminals by posting "WANTED" posters in places where many citizens would see them, such as at the post office where they get their mail. With emails and social media sites rapidly replacing the old "snail mail" as the preferred method of communicating, some police officers have adapted and have moved their "WANTED" posters to social media.

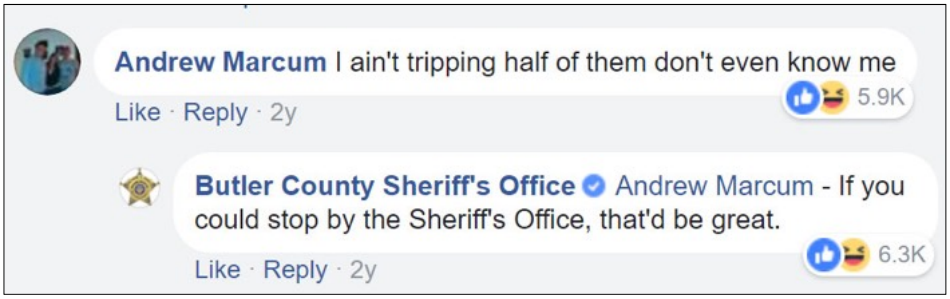
An excellent example of this is the Butler County Sheriff's Office in Hamilton, Ohio. The use their Facebook page (https://sec487.info/9) to showcase wanted criminals. Their local citizens can check the page and provide them tips about the suspects. In the slide above, the sheriff's office is looking for Andrew Dale Marcum because of "numerous warrants," including burglary and abduction. On March 2, 2015 they posted this Facebook entry.

Image from https://sec487.info/9, August 24, 2016.

Butler County Wanted Responses

The suspect replied to the Sheriff's Office post at 2:36pm.

The Sheriff's Office invited the suspect to "stop by".



Butler County Wanted Responses

On March 2 (still) at 2:36 p.m., the suspect replied to the Sheriff's Office stating, "I ain't tripping half of them don't even know me." Three hours later, he received an invitation to "stop by the Sheriff's Office."

Image from https://sec487.info/9, August 24, 2016.

Butler County Captured

The next day, the Sheriff's Office posted that the suspect had turned himself in

A successful use of social media to apprehend a wanted suspect

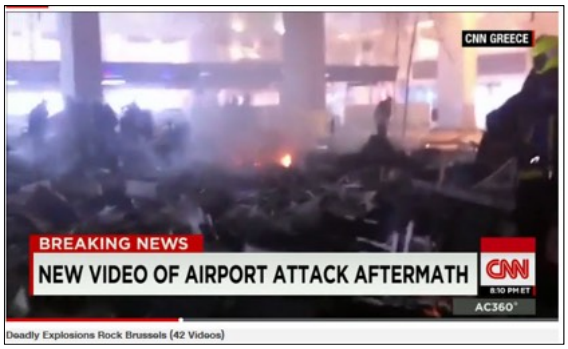
Butler County Captured

The next day, Andrew Marcum turned himself in. In this case, through social media, law enforcement communicated to the specific suspect that they wanted to reach, a task that may have otherwise taken days or longer to achieve. Their social media campaign was a success in engaging their community to find and capture a wanted person.

Image from https://sec487.info/9, August 24, 2016.

Law Enforcement Agencies Use OSINT – Video Surveillance

- Seems like every new electronic device has a camera
- When something happens in the world, people record it
  - Police brutality
  - Dash cams for accident recording
  - Brussels Airport Bombing



Law Enforcement Agencies Use OSINT – Video Surveillance

With the many electronic devices that have cameras and the large number of webcams and other record devices in the public domain, police officers can re-create crimes, get important leads, and view events from a variety of viewpoints. Whether the event is one of a terrorist attack, such as the 2016 bombing of the Brussels airport in Belgium (https://sec487.info/a) or crimes of suspected police brutality, videos of these events make it onto the internet and can be harvested for data. The geolocated social media posts in Snapchat, Facebook, and Periscope can be viewed, and different perspectives of events can be analyzed.

Image https://sec487.info/a, August 27, 2016.

Law Enforcement Agencies Use OSINT – Dark Web Monitoring

- We have heard about dark web marketplaces that sell drugs, guns, and credit cards illegally
- Law enforcement uses OSINT to monitor these web sites and then, sometimes, takes them down

WIRED May 9, 2019

bigger than the Silk Road had ever been. In a well-coordinated, two-pronged attack, the FBI took down Alphabay in July of that year while Dutch police hijacked the second-largest dark-web market, Hansa. That maneuver drove Alphabay's refugees into a trap: The Dutch police had rewritten parts of Hansa's code to de-anonymize users, grab their passwords, and even install beacons on their computers. The double takedown, called Operation Bayonet, was intended not only to ensnare dark-web buyers and sellers but to scare them, too, as the Dutch police's National High Tech Crime Unit told WIRED at the time, creating a deterrent to keep users from migrating to the next dark-web drug bazaar.

Law Enforcement Agencies Use OSINT – Dark Web Monitoring

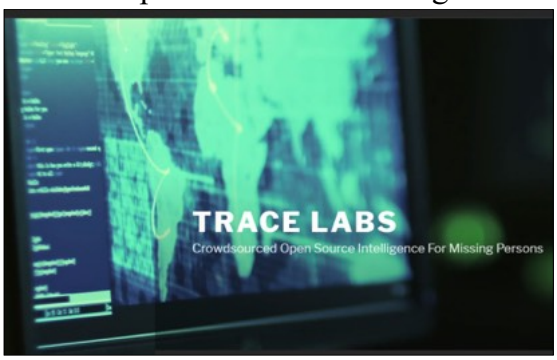
You might have read articles or heard reports of criminals running illegal dark web marketplaces that sell everything from stolen passports and credit card numbers to guns, drugs, and more. Law enforcement knows of these places, too, and use OSINT to monitor the people running and using them. In 2019, the Dutch police's National High Tech Crime Unit monitored, took over, and ran the Hansa dark web marketplace. Using OSINT and then active techniques, they shut down the site, caught many of the buyers on the marketplace, and may have accomplished their goal of creating distrust of dark web marketplaces for those wishing to use them.

Image https://sec487.info/q3, August 9, 2019.

Trace Labs – OSINT CTFs

- Works with regional law enforcement in an area
- Gets missing persons cases < 45 days old
- Creates regional or global Capture The Flags (CTFs) to find information about these missing people
- Hands data to law enforcement for action

https://www.tracelabs.org



Trace Labs – OSINT CTFs

The Trace Labs organization (https://www.tracelabs.org) focuses on assisting law enforcement (LE) teams around the world through their online OSINT Capture The Flag (CTF) platform. They work closely with regional and international LE groups to find reported missing persons cases that are relatively recent. Then, at different OSINT, cyber, and other events, the Trace Labs team allows participants to try their OSINT skills at finding data about the missing people. Their participants find data about where missing people are, who they were last seen with, their friends and relatives, and then submit these bits to the CTF scoring system to get points. Human judges examine the data submitted and determine appropriate point values for each piece sent in. Teams win by submitting the flags with the highest point values and scoring more than the other teams.

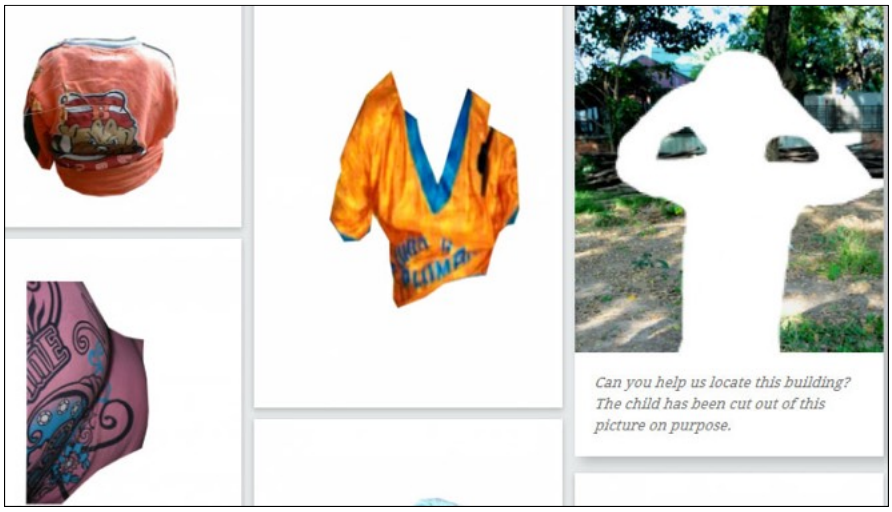
Meanwhile, Trace Labs collects CTF submissions and provides them to the law enforcement teams trying to find the person who is missing. Through the Trace Labs CTFs, several missing persons have been found!

Image https://www.tracelabs.org/, August 9, 2019.

Europol's Stop Child Abuse Site

Like an OSINT challenge for a great cause? Europol posts images associated with child abuse

Find the objects and help them solve cases



Europol's Stop Child Abuse Site

Europol posts[1] images that, once more OSINT work is performed on them, could help solve child abuse cases. Above are some of the images that need to be identified. Whether it is clothing, and Europol needs help identifying where it is sold, or the location of a building, your assistance can help provide law enforcement valuable clues to solve crimes.

Reference:

[1] Image and reference from https://sec487.info/xl, October 3, 2019.

THE BADASS ARMY

Uses OSINT to combat non-consensual pornography

"*A nonprofit organization dedicated to providing support to victims of revenge porn/image abuse, and eradicating the practice through education, advocacy, and legislation.*"[1]

THE BADASS ARMY

Thousands of members of THE BADASS ARMY work to help victims of non-consensual pornography (NCP) by using OSINT, education, and by changing laws. Led by Katelyn Bowden, this nonprofit group uses OSINT to discover where NCP images and videos may be and to unmask who is posting them.

Reference:

[1] Image and reference from https://badassarmy.org/, August 9, 2019.

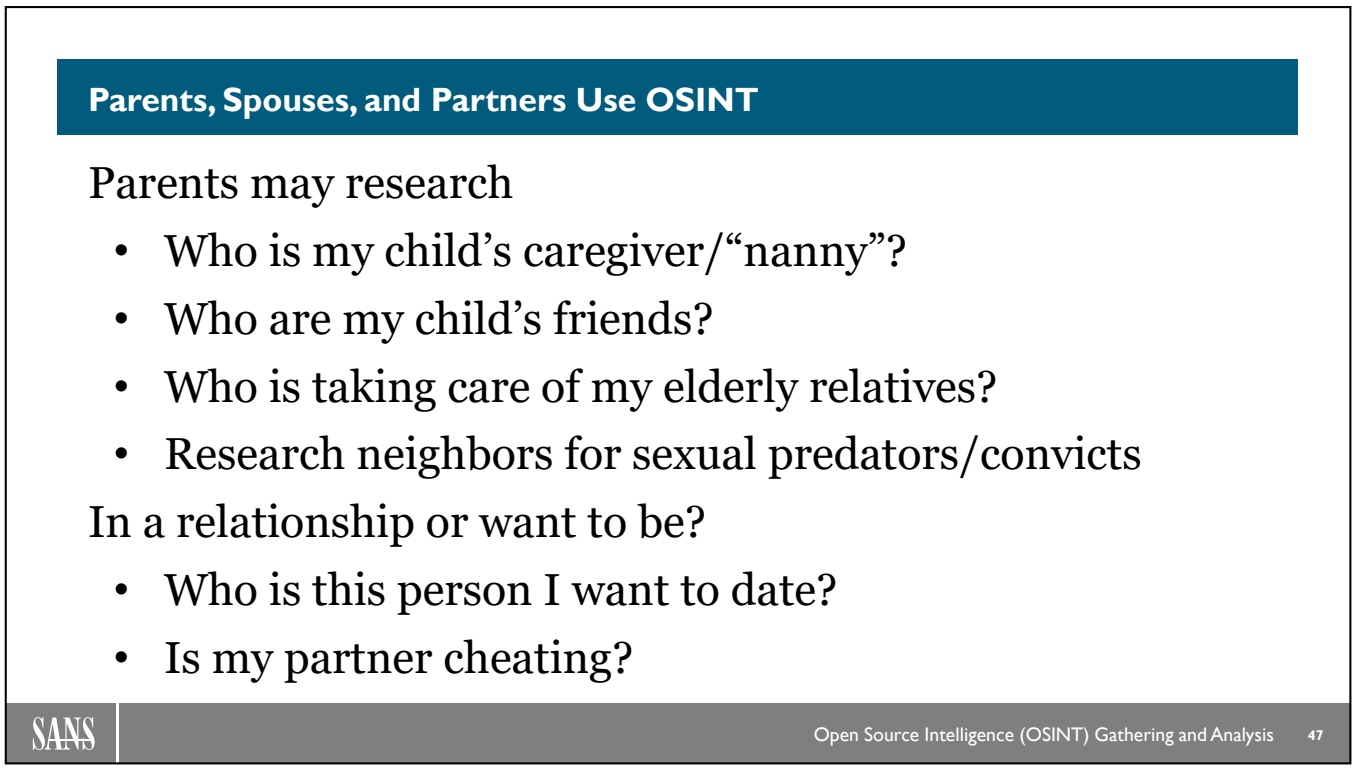
Innocent Lives Foundation

- Chris Hadnagy (@humanhacker) and a team of volunteers doing OSINT
- Searching for online predators
- Bridge between OSINT, Cyber, and Law Enforcement

Innocent Lives Foundation

There comes a time in many peoples' lives when they look for ways to make the world safer for others by using their talents and skills. The acclaimed social engineer Chris Hadnagy (@humanhacker) did that and created a foundation for others to join him in "unmasking online child predators" using their OSINT, cyber security, and investigative skills. More information can be found at https://sec487.info/gn.

Image https://sec487.info/gn, March 26, 2018.



Parents, Spouses, and Partners Use OSINT

Understanding more information about the people who touch our lives and the lives of our families is another reason why we use OSINT and SOCMINT in our daily lives. Some examples of this are:

- Parents may research:
  - The caregivers who work with their children.
  - Their child's social media persona, what they are posting, and who their connections/friends are.
  - The people or organization used to care for elderly relatives.
  - The neighborhood that they live in to discover any sexual predators who may live nearby.
- In an existing relationship? Perhaps OSINT can help determine if a partner is cheating.
- People starting relationships may perform OSINT on their new partner.

For the last scenario about dating, there is a terrific example of this from the American TV show *How I Met Your Mother*, episode "Mystery vs. History" in 2011 (https://sec487.info/ik). In this show, the character Ted meets a woman at a bar. To make the date more like old times when people couldn't research everything about their date online, both of them decide not to use their phones to investigate the other. Unfortunately, Ted's friends use THEIR phones to research his date and find out something that ultimately ruins the evening and the date.

Businesses Use OSINT

- To find competitive intelligence on their rivals
- In the hiring/interview processes
- Looking for malicious insiders
- Finding "missing" employees
- Mergers and acquisitions
- Legal support
- Recruiters and talent sourcers
- To protect their systems and data from attack

Businesses Use OSINT

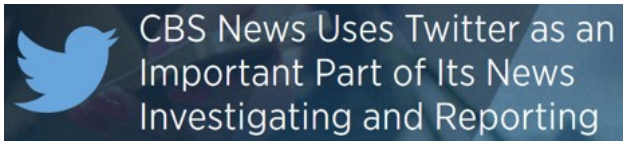
People in the business world leverage OSINT for their purposes as well. This OSINT collection process is usually more formalized than a person who just googles information about their date. Some common business drivers and scenarios where a business may seek OSINT data are:

- To find competitive intelligence on their rivals. Finding out what new products a competitor is working on, who they are hiring and in what capacity, and what prices they are charging their customers can help a rival organization get ahead of their competition.
- Performing searches on job candidates during the hiring process can highlight topics that may be a conflict of interest or may need more clarification.
- Examining what current employees are doing in their "outside of work" lives can show suspicious patterns of behavior, dangerous affiliations, and instances where a company may have a malicious insider.
- Some companies may use OSINT and SOCMINT to find employees who should be at a client site, remote office, or traveling but never arrived at their destination.
- During mergers and acquisitions, leaders of a company that is to be bought may be researched. Results of these searches could yield improper dealings, court issues, and other topics that may be of concern.
- When a company conducts a fraud investigation, OSINT analysts may be called in to provide SOCMINT and other intelligence to discover what is really happening.
- Many of the corporate OSINT investigations that are conducted are to support legal departments or outside entities.
- Recruiters and sources trying to fill open positions use social media and advanced search engine queries to find their targets.
- Cyber defenders use OSINT to find attackers, understand their motives, and examine systems and malware.

Image from https://sec487.info/jc, August 27, 2016.

Media Entities Use OSINT

- Instant reporting of events
- "Eyewitness accounts"
- Researching stories



Solution

Solution CBS News recognized that it was crucial to integrate Twitter into its news reporting. They began by incorporating Dataminr(@dataminr), Tweetdeck(@tweetdeck) and Curator(@Curator) into their news discovery process.

- The downside is that this is under-validated content
  - Lacks analysis
  - Lacks context

Media Entities Use OSINT

In the past, mainstream media companies were the only ones that could broadcast information, images, and videos to wide sections of the population. But now, using a simple tablet or smartphone, anyone has the ability to share this data. For years we have seen television stations report using tweets, Instagram pictures, and YouTube videos. Television stations long recognized the power of crowdsourcing their reporting by using self-reported information from all of us. In 2011, the AdWeek web site (https://sec487.info/c) reported on the process by which many tweets make it onto TV. Twitter's web site promotes that CBS News uses Twitter content to discover and enhance their news reports.

Eyewitness images and videos from natural disasters, sporting events, concerts, and crime scenes provide investigators near-real-time accounts of historical events in the making. Much of this information is under-validated and raw. It provides one angle, one side of a story, and can show a situation from a single person's perspective.

In the 2016 United States Presidential election, Facebook and Twitter sites were used to spread disinformation about topics and people to confuse voters.

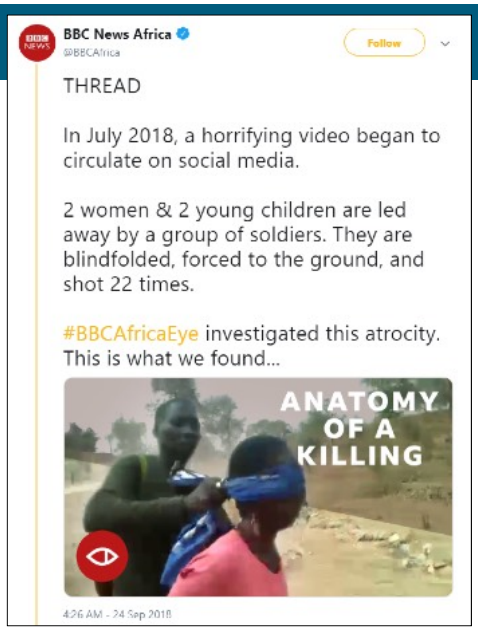
Images from https://sec487.info/ic, August 21, 2018.

BBC Journalists Using OSINT

Many investigative reporters use open source intelligence to gather facts for articles and blogs

One example is the BBC analysis of a Cameroon execution video in 2018

They figured out dates and locations when and where it happened, countering the government's denials



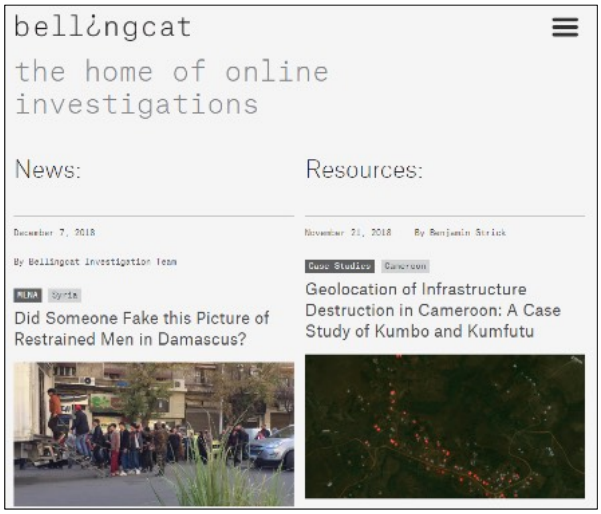
BBC Journalists Using OSINT

Investigative reporters use OSINT in their work to unearth facts. An excellent and horrific example is the work that Benjamin Strick (@BenDoBrown), Aliaume Leroy, and others put in to find the factual details of what happened in a video of a woman and children being executed in Cameroon. At first, the Cameroon government denied that it occurred in their country. Strick and team used Google Earth and satellite imagery to match physical features found in the video (https://sec487.info/iy). Then they matched dates using aerial imagery. Finally, they used social media to find the people that committed the acts.

Image from https://sec487.info/ix, December 11, 2018.

Bellingcat's Online Investigations

"Bellingcat uses open source and social media investigation to investigate a variety of subjects, from Mexican drug lords to conflicts being fought across the world. Bellingcat brings together contributors who specialise in open source and social media investigation, and creates guides and case studies so others may learn to do the same." [1]



Bellingcat's Online Investigations

A fabulous method of learning open source data analysis and research is by examining what others have done. Bellingcat's articles not only show what their reporters and analysts have found, but many times they show how they found their facts. Bellingcat's articles and investigative journalism are deep and heavily researched to ensure accuracy.

Reference:

[1] https://sec487.info/j0, December 11, 2018.

Image from https://www.bellingcat.com/, December 11, 2018.

Intelligence Agencies Use OSINT

- Finding and monitoring terrorist networks
- Attribution of events
  - Cyber – Indicators of Compromise (IOCs)
  - Real-world
- Movement of resources (troops, military equipment)
- Targeting for attacks (cyber and real-world)
- Recruitment of assets

Intelligence Agencies Use OSINT

Intelligence, "intel," or "spy" agencies around the world harvest huge quantities of OSINT daily. They collect, process, analyze, and report on the OSINT happenings of various targets, ranging from people and businesses to adversaries and allies. Since the September 11, 2001 terrorist attacks inside the United States, there has been a spotlight on intelligence agencies and what they can and cannot collect, where they collect it from, and who uses that information. Some of the OSINT collected is used for:

- Finding terrorists and defining their networks. Just as law enforcement uses OSINT to map gang affiliations, intel organizations do the same with terror and extremist groups, tracking their activities and movements through the internet.
- When a cyber or real-world attack occurs, intelligence agencies seek to understand what person or group conducted the attack. OSINT can be useful in this search.
- Governments want to know where adversary military forces are, where they have been, and where they may go next. Aerial photography and social media posts from troops divulge this data. Longer term monitoring of these groups can also be a goal.
- Intel agencies are sometimes engaged to locate targets for attack, be they cyber, human recruitment, or a physical attack. Scouring the open source data provides supporting data for their work.
- The collection of technical and scientific research of companies or governments may be the focus of OSINT activities.

OSINT for "Bad"

- Cyber Attack
  - Phishing, vishing, pretexting, identity theft
  - Objective details, computer systems, network addresses, software
- Theft
- Extortion
- Assault
- Kidnapping
- Organized Crime
- Terrorism
- Stalking, harassment, bullying, criminal mischief

OSINT for "Bad"

OSINT by itself is neither good nor bad; it just is. What the collector does with the data can then be described as good or bad. Since we covered the many ways that people use OSINT for helpful and ethically good reasons, let us now touch on how people with bad intentions might use this same data and these same techniques.

- **Cyber Attack** – Reconnaissance is the hallmark of an excellent cyber attack. Performing OSINT on targets and potential victims can help attackers gain access to online accounts (as happened to Sarah Palin in 2008[1]), learn patterns of their victims, and gather important information for later use in future phishing, vishing, pretexting, and identity theft attacks. Attackers can find information on the internet about what computer systems their targets use, how many there are, how they are configured, and potential weaknesses.
- **Theft** – Online activity posting and geolocation features allow attackers to understand where victims are and where they are not. There are documented attacks against the property of users using activity-tracking software such as Strava and MapMyRide.[2]
- **Extortion** – Using information gathered online, attackers can discover sensitive, private information about victims and try to use that to influence their victims. After the Ashley Madison and Adult Friend Finder compromises, there was an increase in such attacks on the people whose information was found in the dump records.[3]
- **Assault/Kidnapping** – With the vast amount of geolocation-tagged "check-in" data that users publish to the internet, attackers can find patterns in their behaviors and use these observations to meet up with their victims and hurt or kidnap them.
- **Organized Crime** – Finding targets that may be open to taking bribes or easily coerced is just one of the reasons organized crime syndicates might use OSINT.
- **Terrorism** – From scoping out traffic patterns and digital imagery about a physical location to understanding when certain events will occur to recruitment opportunities, terrorists use OSINT to further their goals.

- **Stalking, harassment, bullying, and criminal mischief** – We will examine many methods for tracking the activities of people and discovering sometimes private information about them. Stalkers and people who wish to hurt others can leverage this same information to find vulnerabilities in their victims and exploit them.

References:

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

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

[3] https://sec487.info/g

Strava Equipment Theft

Police thought thieves were looking at peoples' exercise habits

Then, using OSINT, they would find their homes and steal their expensive bikes

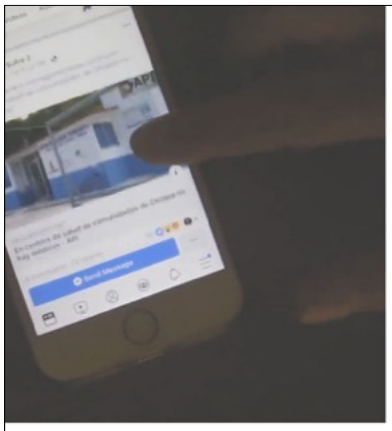
Strava Equipment Theft

In the United Kingdom in 2015, police thought thieves were using OSINT on users of exercise applications to find people with expensive equipment. They would then find where the targets lived, break into their garages, and steal expensive bikes. More information can be found https://sec487.info/f.

Cartels Use Social Media

We are not the only people who use OSINT to find people

In 2019, *USA Today* published a report that noted Mexican cartels were using social media to extort money from the family members of their hostages



'We're going to find you.' Mexican cartels turn social media into tools for extortion, threats and violence

By Rebecca Plevin and Omar Ornelas
Updated 7:34 p.m. PST Feb. 28, 2019

Hit lists published on platforms like Facebook and WhatsApp drive people to flee, but even once they're in the U.S. they continue to be stalked.

Cartels Use Social Media

Imagine going about your day at your job and receiving a text message or social media direct message stating that a drug cartel in Mexico had one of your relatives and was going to harm them if you didn't send money immediately. Sounds like a phishing ploy, yes? Well, in 2019, *USA Today* published an article (https://sec487.info/q2) about how this nightmare of a situation was happening to people across the United States. Their family members who were visiting or living in Mexico were being taken hostage, and the cartels, using OSINT, would find the hostage's family members and demand money in return for their safe return.

Image from https://sec487.info/q2, August 9, 2019.

What Do Our Customers Want?

We have seen the realm of what people "can" use OSINT for, both for bad and good.

While our customers may think they know what they want, we need to educate them about our work, what is possible and what is not.

Let's explore some of the questions our customers may ask.

What Do Our Customers Want?

Since we better understand why people use OSINT, we now need to look at OSINT from our customers' perspectives. They see television shows and movies where amazing feats are accomplished in seconds using computers and simple queries of the internet. Yes, this sometimes can occur in our assessments. Before we begin the investigation, we need to understand our customers' motivations for having us perform an OSINT investigation. We will need to educate our customers about what we can and—due to time constraints, ethical and legal limitations, and other issues—what we cannot do within the scope of the assessment. Let's look at some of these issues and see how we might explain them to our customers.

Assessment Goals from Our Customers

Do they have a general question or just want research?

What is a successful assessment for your client?

What is the motive of the client?

- Solving a crime?
- Finding a missing person?
- Are they a stalker and you are going to help them find their next victim?

Consider checking with the client's and your legal counsel

Assessment Goals from Our Customers

There is usually a reason for your customers to request an OSINT investigation. During your pre-engagement talk, you need to ensure that you understand what questions they want answered through your work.

- Sometimes your customers will want general reconnaissance performed in preparation for some other event (such as a merger with another company). In other instances, there is a specific question that your customers will want to know about. Examples of this could be:
  - "Is my spouse cheating on me?"
  - "Is the target person really injured from their car accident?"
  - "Who is this nanny that I'm hiring?"
  - "Who are the friends of the target and what religious/socio-political views does that group of people share?"
- You need to inquire what a successful engagement might look like for your customers. A good example here is if a person asks you to find out if their spouse is cheating on them, they may consider a negative result (i.e., no cheating evidence was discovered) just as valid an outcome as a positive one (i.e., yes, there is evidence of unfaithfulness). Find out what your customer is looking for and ensure your investigation addresses those issues.
- As we have seen, OSINT can be used for good and for evil. Ask your customer about the "whys" of the assessment. Why are they looking to have this investigation performed? Why are they concerned or not concerned about the target? This can lead to valuable starting information for you. Ensure you are comfortable with the customer and what they are asking for. It would be unfortunate if you found a person your customer was looking for and reported their location, only to see the next week that the target had been assaulted or worse.

When applicable, you may wish to consult legal counsel before beginning some less-standard investigations. Staying on the correct side of the laws of your locality, state, and country is important for you and your business.

Examine Techniques and Outputs

What are the client's requirements for...

- Report output?
- Data storage during and after the assessment?
- Covertness during the investigation?
  - Interaction with target or no? If so, ensure legal input before starting
  - Looking at relatives to get to target (relatives' profiles may be less secure and have data of target)

Examine Techniques and Outputs

With a better idea of why your customer wants the work completed, you may still have more procedural questions that need to be answered. What kind of output does your customer want? Some customers just want a simple answer to their question, and others may want additional details or a full report. Some companies require that a report be produced for each OSINT investigation, unless a client signs a waiver stating they don't need one. Ensure that you understand these reporting requirements prior to the engagement.

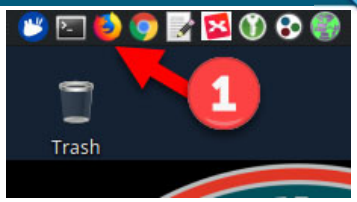
Does your customer have specific guidelines for how you need to store the data you gather during the assessment? Will you need to use encryption when the data is on your computer? What does your customer want you to do with your notes, report, and supporting evidence after the assessment is completed? Dispose of them? Retain them for a certain length of time in case the customer needs them? Ask the questions.

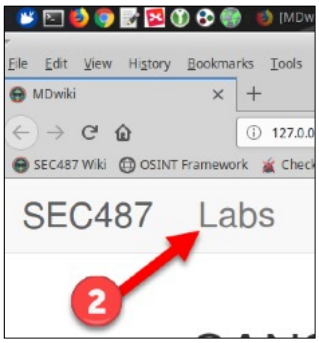
Another topic that should be raised is what level of covertness does your customer want you to take during your work? Do you need to become an OSINT ninja and not leave any tracks of the investigative work, or will they allow you to interact with the target? Remember to talk to your legal team and ensure that your customer does the same before you interact with the target. While we are discussing interactions, it is common that the friends and family members of targets sometimes have less-secure profiles. Does your customer allow you to research those close to your target? What about connecting with them on social media? Ask.

SEC487 Workbook

Please visit the course electronic workbook in the 487 virtual machine

1. Launch Firefox
2. Click the "Labs" link
3. Navigate to the 487.1 Lab named "**0 - Setup**"





This page intentionally left blank.

SEC487 Workbook

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

"Setup"

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. Goals of OSINT Collection
4. **Diving into the Collecting**
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. Leveraging Search Engines

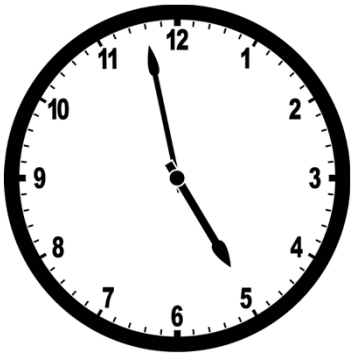
This page intentionally left blank.

A Fictional Example

It is 4:58 p.m. Two minutes until the workday ends.

It is the last day of the work week.

Mr. Vizzini, the OSINT person in the office, is wrapping up one final email and then it is the weekend!



A Fictional Example

Some people's actions serve as an example of how NOT to do things. Let us walk through a fictional example of how an OSINT investigation could be executed. While the events in this example are not real, think to yourself about how many times pieces of this may have happened to you or "a friend." And with that, we will begin the story:

It is 4:58 p.m. on the last work day of the week, and there are just two more minutes until the end of the workday. The weekend is almost here. Mentally, Mr. Vizzini is already thinking about his weekend plans. He has just one more email to finish writing and then he is done.

Image from https://sec487.info/jb, August 14, 2016.

Oh Nooo!

And THAT is when his boss comes over.

"Yeaahhh. So, the company has been hacked."

"Attacker stole all our email."

“You do that OSINT stuff.”

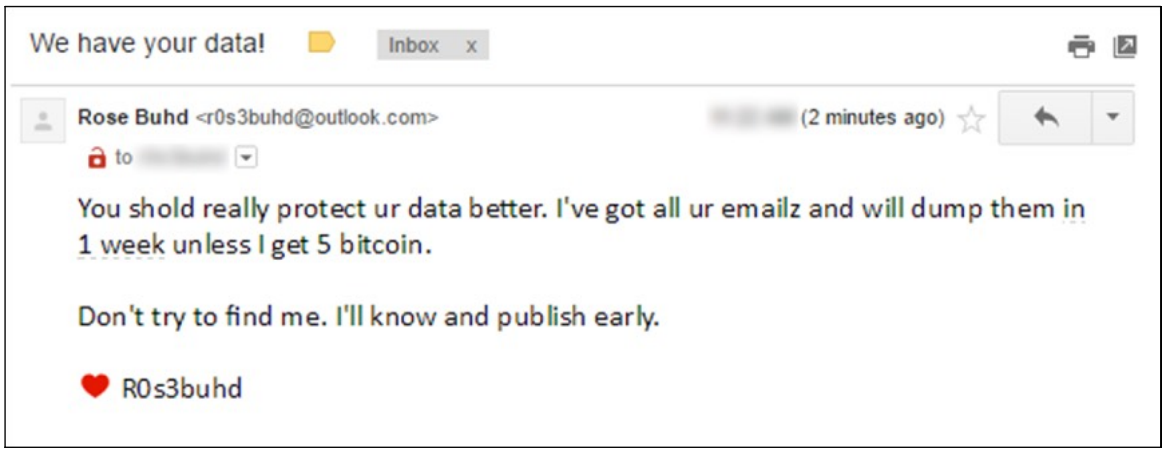
"We need you, Vizzini, to find the attacker."

Oh Nooo!

It is just then that Vizzini’s boss goes over to his desk and tells him the bad news: "The company has been hacked, and an attacker has stolen all the company’s emails." He needs Vizzini to find the attacker using his OSINT techniques.

Image from https://sec487.info/ja. August 14, 2016.

We Have an Email to Start



We Have an Email to Start

Vizzini's boss shows him an email from the attacker whose name appears to be "R0s3buhd."

The email taunts the company about poor security and mentions that, if the company doesn't pay R0s3buhd 5 bitcoin in 1 week, she will dump all the emails to a public place. She adds one final warning: "Don't try to find me. I'll know and publish early."

She signed the email "R0s3buhd," and even put a heart emoji. This must be the age of the kinder, gentler attacker.

The Email Has Good Data

The Email Has Good Data

Some people falsify their email address when sending these types of email, but not this attacker. That "r0s3buhd@outlook.com" could be a great starting point for an OSINT investigation. There is also the unique attacker name, "R0s3buhd," that could yield some valuable search results.

Vizzini Gets to Work

With his boss looking over his shoulder, Vizzini googles it:

Vizzini Gets to Work

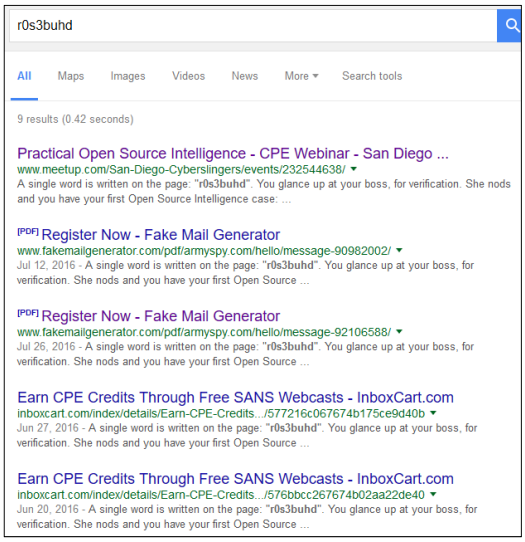
With his boss looking over his shoulder and urging him to start the OSINT investigation, Vizzini launches a web browser, loads up the Google search engine page, and enters the "r0s3buhd" name into the search field. After clicking the Google Search button, both Vizzini and his boss eagerly await the results.

Google Results

The results are for a webcast, a meetup, and a SANS event.

False positives. Darn!

So, Vizzini checks social media sites.



Google Results

Google provides very few results, and most are for a SANS Institute webcast and a "meetup" event. Those aren't the attacker. Vizzini moves on to checking social media sites.

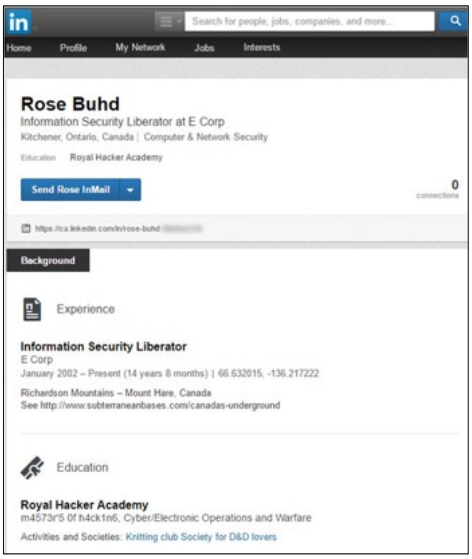
Vizzini Actually Gets a Hit!

There is a LinkedIn account for "Rose Buhd"

- Information Security Liberator?
- E Corp?
- m4573r'5 0f h4ck1n6 (Master's of Hacking) degree?

Uh oh.

This could be bad.

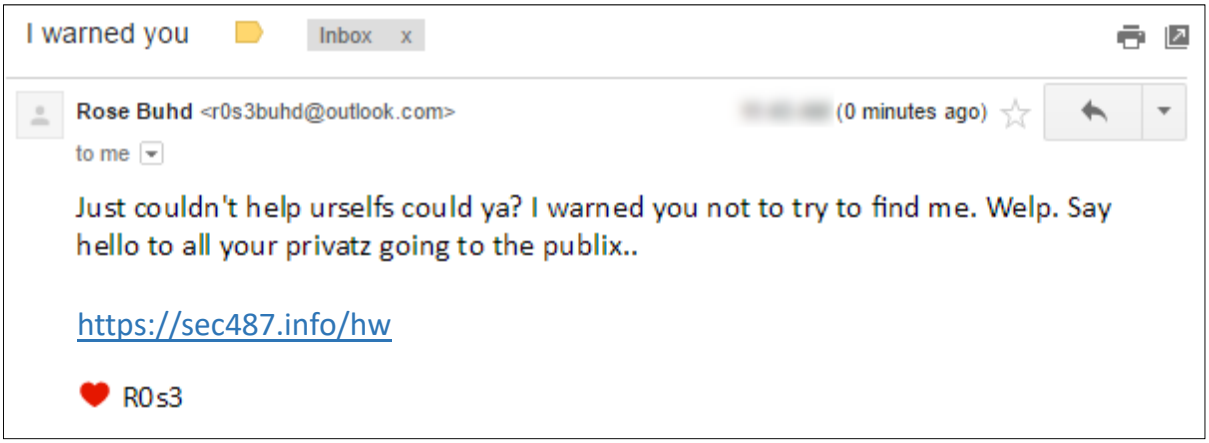


Vizzini Actually Gets a Hit!

The social media search paid off, and Vizzini has the profile for a Rose Buhd from Canada. That could be the attacker. Searching further down the LinkedIn profile page, Vizzini notices that Ms. Buhd works as an "Information Security Liberator" at E Corp? That sounds suspicious and, wait, isn't E Corp from the *Mr. Robot* hacking TV show? Vizzini is not getting a good feeling about this. Further down on the page he notices that Ms. Buhd has an "m4573r'5 0f h4ck1n6" (Master's of Hacking) degree. He didn't even realize that there was such a thing. This does seem like it could be the attacker.

Just then, Vizzini's boss's phone dings with a new email.

The New Email



The New Email

The email is from the company's attacker. She noticed that someone was doing OSINT on her and is going to release Vizzini's company's emails now! She even provided the https://sec487.info/hw short URL to tell them where their private emails can be found.

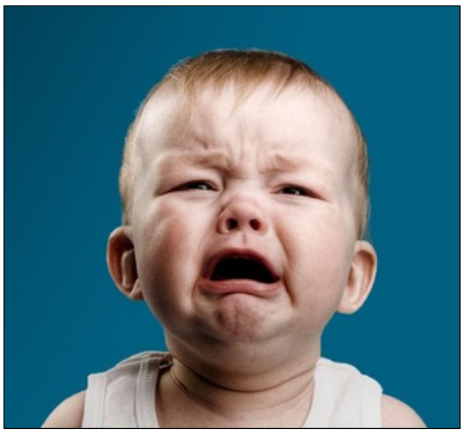
Vizzini's Postmortem

The company's data is gone

But out of failure comes opportunity

Ever do a cyber event postmortem?

Figure out what happened to improve future actions



Vizzini's Postmortem

With the company's email published on the internet, Vizzini and his boss conduct a postmortem, after-action review of what went wrong. These events can help improve process and discover process flaws.

Image from https://sec487.info/j8, August 17, 2016.

What Happened?

- ✓ Vizzini googled a name
  - ✓ No significant/helpful results were found
- ✓ Queried a social media site (LinkedIn)
  - ✓ Discovered a profile of interest
  - ✓ Visited the profile page
- ✓ Received new email from attacker

What Happened?

Let us see if we can better understand what happened through analyzing the main steps that Vizzini followed in his brief and unsuccessful OSINT investigation.

1. Vizzini Googled a name and found some results that appeared to be something other than his attacker. It is common to find false-positive results in assessments.
2. After that, Vizzini used LinkedIn's search feature to look for a person with a similar name. Translating the name "r0s3buhd" from leet speak[1] could yield the name "Rose Buhd." Luckily, that was a valid account on the LinkedIn website and Vizzini retrieved her profile page.
3. That was when the attacker emailed Vizzini's boss that she knew they had been performing OSINT on her and, because of it, she was going to release their emails to the public.

Reference:

[1] For more information on leet speak, visit https://sec487.info/h.

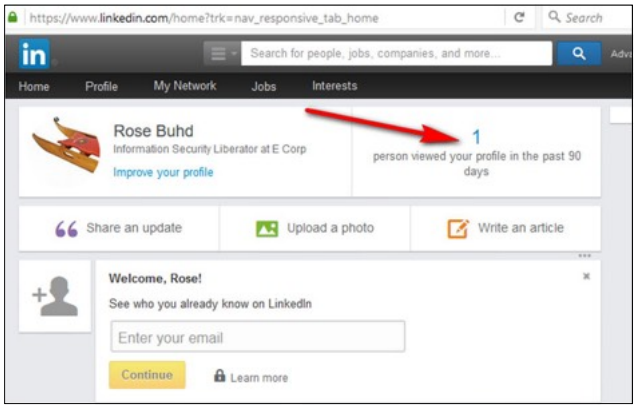
Image from https://sec487.info/j7, August 17, 2016.

It Was Probably LinkedIn

When we visit another's profile on LinkedIn and are logged in, it tells them who visited

R0s3buhd saw that Vizzini visited her profile

Visiting his profile showed he worked for the victim company



It Was Probably LinkedIn

While there are other places where Vizzini's work could have tipped off R0s3buhd that someone was researching her, in this situation, it was most likely the LinkedIn page viewing that did it. Vizzini didn't know that visiting target profile pages on social media sites could allow the target to see your activity. Vizzini was logged into his profile on LinkedIn when performing his searches and viewing Rose Buhd's profile. See the arrow pointing to the "1 person viewed your profile" section of the page? That was Vizzini's visit.

Vizzini's profile page probably noted the company that he currently worked at, and Rose put the facts together and knew that the victim company had researched her.

How Do We Do It Better?

1. Use a reliable process
2. Gather requirements before working
3. Be aware of your network traffic and what it gives away
4. Understand and properly configure your system
5. Don't use your personal accounts for your work
6. Document everything

How Do We Do It Better?

In this course, the class will start at the very beginning. Learning from Vizzini's errors, we can put processes in place to make an OSINT investigation more covert. We will be learning a process in class, and you can adapt it to fit your OSINT work and your customers' needs. As analysts, we will gain a complete understanding of what our customers want us to find and to report. We will also dive into topics such as "how to set up a system for OSINT" and "network issues that could affect your research" to ensure that each student appreciates some of the complex decisions that will need to be settled before an assessment. Not using personal accounts in OSINT research is about insulating the analyst from the target(s) and conducting a safer investigation. Finally, we will need to document everything, as we may not know what is relevant and important in an assessment until we have collected enough data.

SEC487 Workbook

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

"OSINTing People Challenge"

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. Goals of OSINT Collection
4. Diving into the Collecting
5. **Taking Excellent Notes**
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. Leveraging Search Engines

This page intentionally left blank.

Note-Taking for OSINT

- In some industries, the end result is all that matters
  - A thing is made – a pencil
  - An idea, strategy, or slogan is conceived
    - "The milk chocolate melts in your mouth, not in your hand."
- In OSINT, the journey taken to get there is sometimes key
- Detailed documentation may be crucial to your customer, depending upon "why" there is an investigation

Note-Taking for OSINT

When a factory makes a pencil, that pencil is the lasting product that the customer uses. Very few people care about how a pencil is made[1] and the journey that the cedar or graphite took to get to the factory. We just want to use the pencil. Same thing with marketing and advertising content. Most people understand what we are talking about when we say, "the milk chocolate melts in your mouth, not in your hand" (the answer is M&M candies), but few understand how that slogan came into being.

With OSINT, sometimes how we get to a result may be just as important, if not more than the final result. In order to be able to tell that story of how you found out that a target was friends with a person of interest, you will need to document what searches you perform to reach that result. You may not need to provide this to your customer, but you should definitely have it for your own records.

Reference:

[1] If you are one of the people who cares, here is "How A Pencil Is Made" - https://sec487.info/j.

Why Are Notes Important? (1)

- Lasting impression on your customer
- Publish the results on a blog or paper
- Document the work for the future
  - May need to follow up on your previous work
- Support of law enforcement, intelligence customers, legal entities
  - These may have their own requirements for documentation and process

Why Are Notes Important? (1)

Your notes and your reports (which should be the polished output from our notes) are valuable to you and your customers. They can demonstrate your professionalism and high degree of customer service by giving the customer just what they asked for (or by exceeding your customer's expectations!). They are the lasting impression that your customer has for the work you performed. I had a vendor provide my company a templated report that had little value. It had typographical and grammatical errors, and it made me feel terrible about the money we wasted on its creation. That company never won another contract with our organization. Make the report something you and your company are proud of and something that delivers value to your customer.

Sometimes our OSINT may be used as reference content in a blog or news article. Having the details of what content was discovered on which web site and on what day could be important to the overall story being told.

When we deliver our reports, many times that is the end of our work for a specific customer. Or perhaps that task of researching Prince Humperdinck comes to a close and you now will perform an investigation into your next target. In three months or six months or even a year, your customer may let you know that they are taking Prince Humperdinck to court and will be using your OSINT report as evidence. Maybe they asked for your notes when you delivered the report, and they continued the work without you, or perhaps they want you to pick up your old work and continue it after a year has passed. Perhaps they are going to call you to testify at the legal proceedings. With great notes, you can be more successful.

There may be investigations that you execute that support law enforcement, an intelligence customer, or a legal team. This work may require a higher level of documentation so that your work can be used in the judicial system or to support intelligence activities.

Why Are Notes Important? (2)

Others may be using your work as a starting point

- Incident responders
- Malware analysts
- Law enforcement
- Another OSINT analyst
- Legal team

The target (or someone/something else) may alter or delete the records after you find them

- Your notes and copies of pictures/documents need to be solid

Improving your craft through self-analysis and reviewing old cases

Why Are Notes Important? (2)

For some customers, our work represents a starting point. We collect as much information as we can, analyze and report it, and then they take up the investigation where we left it. Providing your customer with your results and with avenues of inquiry that did not pan out will save them time and effort.

Let's face it: in today's world, people can easily learn how to reduce or remove most of their OSINT footprint. Just DuckDuckGo "how to protect my privacy online." Depending upon the web site, your targets may be able to alter the data you are collecting. Information that is there today could be different or gone tomorrow. We document to preserve data.

We also document our work because sometimes we write reports when we are tired or distracted. Sometimes certain content may not make sense, or our customer requests detailed answers. It is easy to get answers when you have performed solid data collection and documentation. As a penetration tester, I can tell you that there have been many times when I have had to go back and consult my notes during and after assessments. Same thing happens in OSINT investigations.

Finally, taking good notes allows you to get better. Review old cases and find weaknesses in your craft. Then you can focus on learning and developing.

How to Document

- The need to document is clear
- The level of documentation is usually driven by your customer's needs and your industry
- Creating and using a repeatable process is key
- Tools can assist us, but we need to understand their uses
  - What are the tool's strengths and weaknesses?
  - Does it need to be used in combination with other tools?
  - How easy/hard is it to use?

How to Document

We have established that we must preserve our thoughts, observations, and analyses by taking excellent notes. The question now is, how detailed must those notes be? Do you need to record every web page that you visited with the corresponding date and time of the web traffic, your IP (Internet Protocol) address, and your user agent string, or is it more appropriate to scale back what you document? These questions will largely be answered by your customer's needs and the requirements of your company. While some customers may only require that your investigation address their requirements, your company may decide that, since you may not know where this data will end up in six months or a year, taking more-detailed notes is more acceptable.

Regardless of the level of documentation you are compelled to create and maintain, you will need to use a repeatable process so that each assessment you (or your team) perform(s) maintains that level of appropriate documentation. As in other areas of our lives, we have several pieces of computer software that will assist us in our documentation tasks. As analysts, we need to understand what each tool's strengths are and when to use it. Some tools require more effort or have different system requirements to run, and that may affect when you decide to use them.

Record Everything

"Record everything" should be the default

- Scale back depending upon time, money, and mission

Content to be recorded:

- Posted images, tool screenshots, pictures of web pages
- URLs
- Times and dates
- Use a sock puppet account? Professional API? Google?
- Use a web browser or a certain tool to find data?

Record Everything

If you receive no direction from your client or your company about what to record during an investigation, record everything you can without it being burdensome. Your primary task is to meet (or exceed!) your customer's requirements so your documentation tasks should not interfere with those goals. You can always scale back or tailor your level of documentation if it becomes too cumbersome.

Now the question is, what might you need to record in your notes? While the list below is not exhaustive, it should provide you with ideas of what you could record.

- Posted images, screenshots of web or other content, pictures of web pages.
- URLs that you visited, especially ones that had important or useful content.
- The dates and times you visited web pages of interest and ran tools. Having a timeline of the actions you took during your investigation can be helpful if you ever are required to work with law enforcement or legal entities.
- How you found and accessed the information. What method did you use to discover what you are recording? Did you use version 1.23 of a free tool or use a sock puppet social media account to view content that would normally not be available to the public?
- You may also not wish to divulge the sock puppet account names that you used to gather the data. This may "burn" the account, rendering it useless for future activities.

Types of Documentation Tools

Visualizers

- Good for analyzing relationships between objects
- Analyzing large data sets
- Can include automation

Note-taking Apps

- Dedicated to recording (manually) notes on your work

Documenting Apps

- Meant for recording your work for you
- Trail of your work

Word Processors/Text Editors

- Report writing
- General documentation

Types of Documentation Tools
There are four main types of software that can be used to document your investigation: visualizers, note-taking applications, documentation applications, and word processors/text editors. Let's take a brief look at how an OSINT investigator may employ these tools.

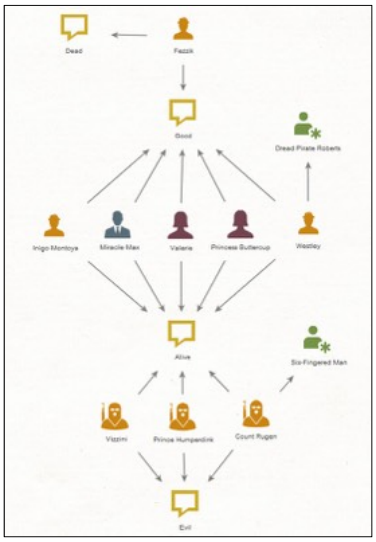
- **Visualizers** take data sets and create a visual display of the information. The goal with this class of tool is to examine the relationship between the different pieces of data in the set. If the data set is a set of all the "friends" that a target is connected to on Facebook, visualizers may link each "friend" to the target and then allow the analyst to check each of those friends to see who they are friends with. The tool will then plot those relationships. This can make groups of friends, gang members, and family relationships easier to see.
- **Note-taking applications** help analysts by allowing us to jot down interesting pieces of information and the steps we followed in our assessment. Many people use these types of apps in their daily lives for taking notes in school, at work, or in other places.
- **Documenting applications** focus on creating an automated timeline of what you have browsed and searched for within your web browser. This class of tool can greatly decrease the burden of timeline creation as well as recording dates and times and where you visit during your work.
- **Word processors and text editors** can also be helpful software. Many analysts write up formal reports after an assessment is complete. This is the lasting content that your customer receives. Just as with the note-taking apps, we can use word processors for general documentation, including timeline creation, note-taking, and embedding screenshots of tool output or interesting data.

While not a documentation tool, there is another method to record your work, and that is with network packet capture software. This class of tool commonly records all the computer network traffic coming from and going to a computer (or computers) and stores it in a PCAP (Packet Capture) file format for later use. The benefit of this software is that you have not only the application layer of what websites your browser visited but you also capture other traffic that came from your tools, DNS (Domain Name System) queries, and other traffic. If your company is truly looking to record EVERYTHING that was sent or received from your computer(s) during an assessment, this is a good method. The downside is that it takes a little more technical skill to retrieve the data from a PCAP file than from Microsoft Word.

Visualizer and Analysis Applications

Examples include:

- i2 Analyst Notebook ($$$)
- Tableau ($$$)
- Maltego (Free to $$$)
- Maltego Casefile (Free)
- Gephi (Free)



Visualizer and Analysis Applications
There are several excellent tools within the visualizer class of tools:

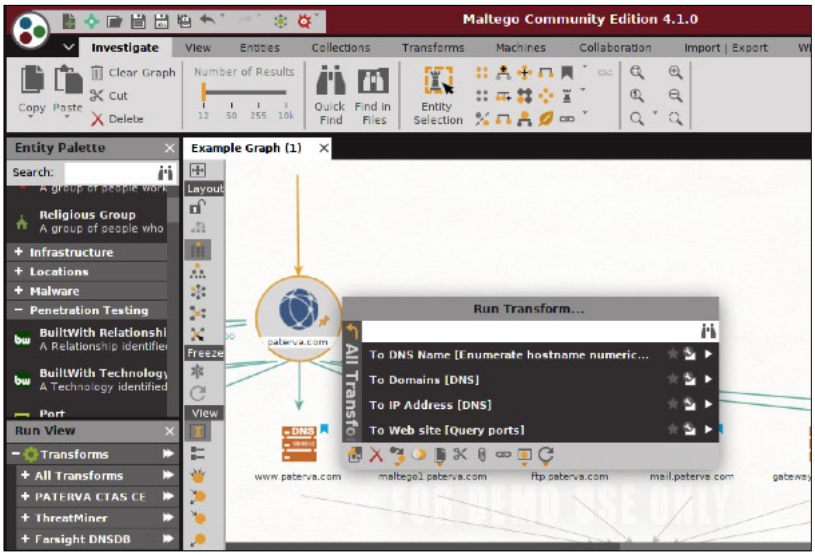
- **IBM's Analyst Notebook** (https://sec487.info/ig) – This software is expensive and used frequently within government intelligence organizations such as the United States CIA (Central Intelligence Agency) and NSA (National Security Agency). It provides not only visualization of data sets but through the creation of objects, such as a person of interest or cell phone number, analysts can create visual timelines, relationships between devices and people and organizations, and other connections. Learning the basics of this tool is not difficult, and analysts should be able to begin using it for basic assessments within several hours. The downside of this application is that each license can be very expensive. A great blog post about using the app can be found at https://sec487.info/3g.
- **Tableau** (https://www.tableau.com/) – Tableau excels at allowing analysts to import and manipulate large sets of related (or unrelated) data. Here we have a tool that can beautifully display trends, relationships, and assist with analysis of a variety of data sets. The Tableau web site has many free tutorials to help users learn the tool quickly. Two challenges with Tableau are that there is a significant learning curve to become proficient at using the application. Analysts will most likely need to spend several days (or longer) learning how to leverage Tableau's features. Also, similar to Analyst Notebook, Tableau can be expensive to buy and use.
- **Maltego** (https://sec487.info/j6) – This visualization application is an excellent method for performing similar tasks to Analyst Notebook and Tableau. Maltego can also use "transforms" to take one data object and change it into another object. Say, for example, you have found your target's Facebook username is "PrincessButtercup." Maltego would create an object for this. By right-clicking on that object, you can ask Maltego to try to find PrincessButtercup's real name. If it is found, an object for her real name would be created, and you could run transforms on that to find email addresses for the name or other information. It should be noted that, unless you use your own personal transform server, Maltego will use public servers that other people in the world maintain and possibly monitor. So, if you are concerned about privacy, anonymity, and your customer's information, you may not wish to leverage Maltego's transforms.

- **Casefile** (https://sec487.info/j6) – Casefile is Maltego but without the transform capabilities. It is also free! For documentation purposes and analysis of large data sets, it can be useful and is easy to use. Compare the different versions of Maltego at https://sec487.info/hl.
- **Gephi** (https://gephi.org/) - Also free, Gephi is an Open Graph visualization platform that can be used to highlight relationships in data points using a nodes and edges method.

Maltego Use

Maltego uses transforms to search on one type of data to get another

Can also use it for visualizations and relationship viewing

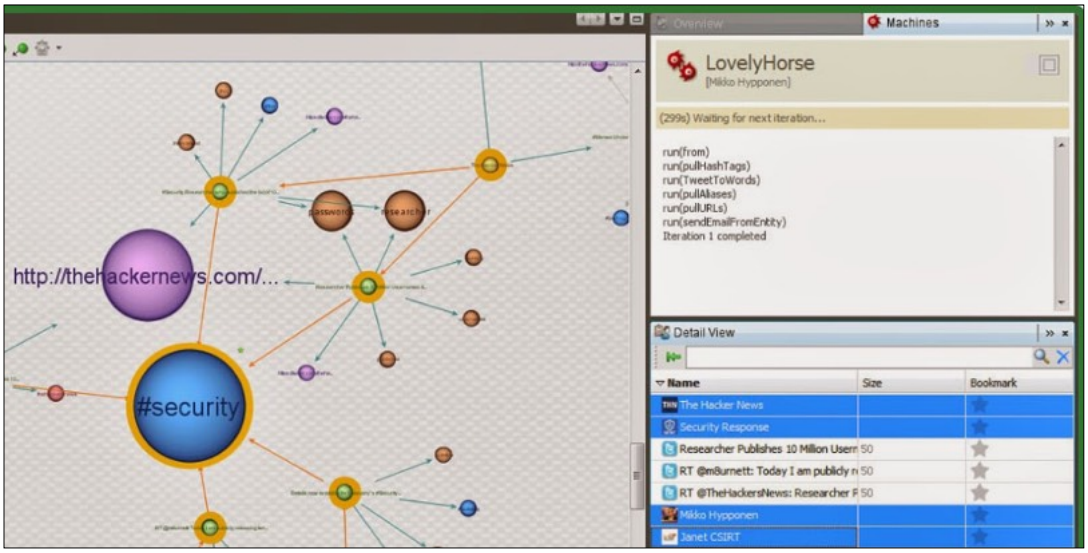


Maltego Use

Maltego is several tools wrapped into a single product. You can either manually or programmatically add objects to the graph (the area where the data is). Each object has certain transformations that can be performed on it based upon what type of data it is. For instance, an email address object can be transformed into a Facebook account, or a domain name can be transformed into the IP address(es) that it resolves to.

As mentioned previously, Maltego has several versions. The free Casefile version has no transformations but does contain the software's visualization features. You can import a CSV into Maltego and then reorient the nodes to best view their connectedness. The other versions of Maltego allow the use of transformations, but some of those transforms require API keys that you might need to purchase.

Maltego Example: Social Network Monitoring



Maltego Example: Social Network Monitoring

The Paterva web site, which is the home of Maltego, hosts some example images that demonstrate the functions and look of Maltego. Above is part of the image from "Social Network Monitoring - keywords, links and hashtags shared by popular security based Twitter users." In it, we can see larger nodes for topics, hashtags, and links that were tweeted more often by certain information security people. The right side of the image shows the details of what accounts are being monitored and what content they are tweeting.

While this image shows social network accounts and topics, Maltego can retrieve data about IP addresses, domain names, and much more.

Image from https://sec487.info/j2, December 11, 2018. (URL no longer active)

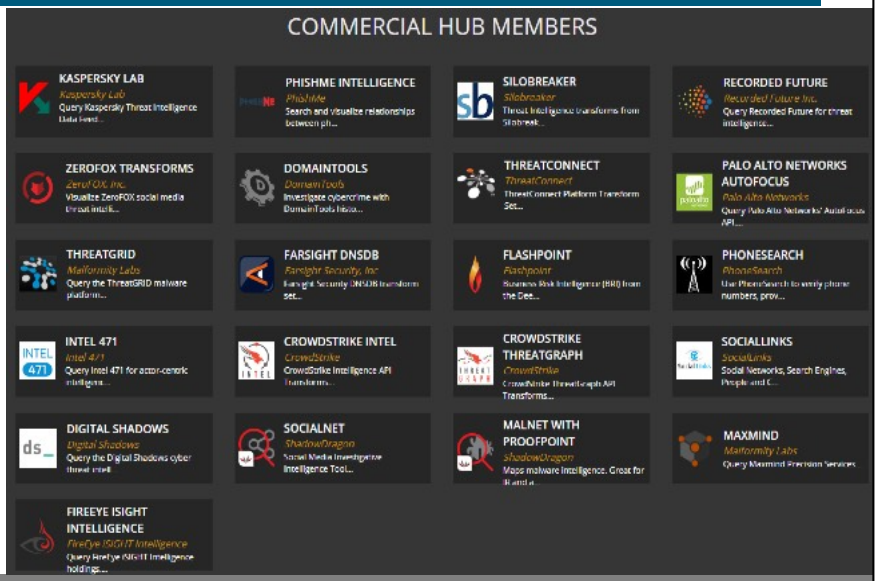
Maltego - Additional Transforms

Maltego comes with basic transforms

Build your own

Download from others

Buy transform packs from the commercial hub



Maltego - Additional Transforms

Aside from the free Casefile tool, all the other versions of Maltego come with basic transforms to take one set of objects and look them up on sites on the internet to retrieve other data about them. You can write your own transforms using the excellent documentation at https://sec487.info/hi, share and download transforms from the internet, and purchase transform packs from suppliers in the Commercial Hub (https://sec487.info/hj). Some of these packs may cost thousands or tens of thousands of US dollars per year, but they may get you access to the data you need more quickly than performing these tasks manually. They also may come with special access to data repositories that, without the transform pack, you do not have access to.

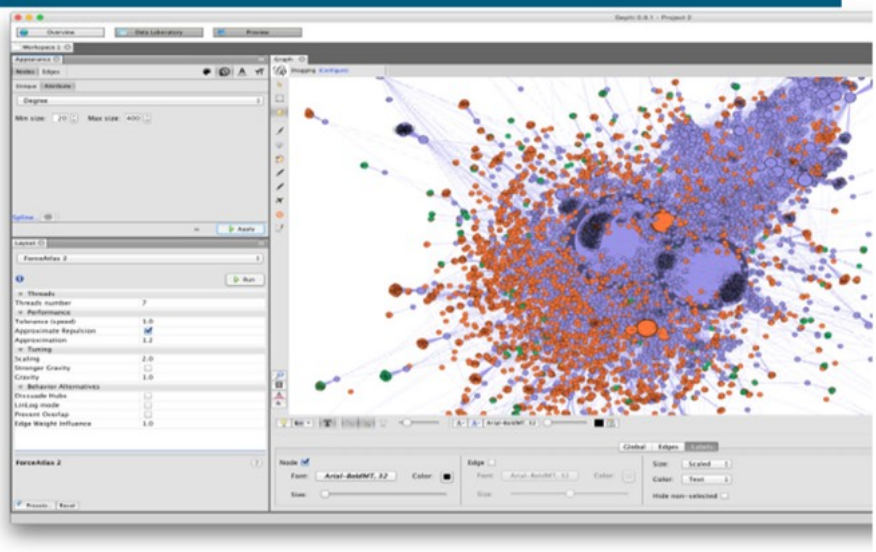
Image from https://sec487.info/hj, March 31, 2018.

Gephi

For large data sets, the free Gephi tool helps visualize relationships between nodes

Data science tool with a steep learning curve

Works on Windows, Mac, Linux



Gephi

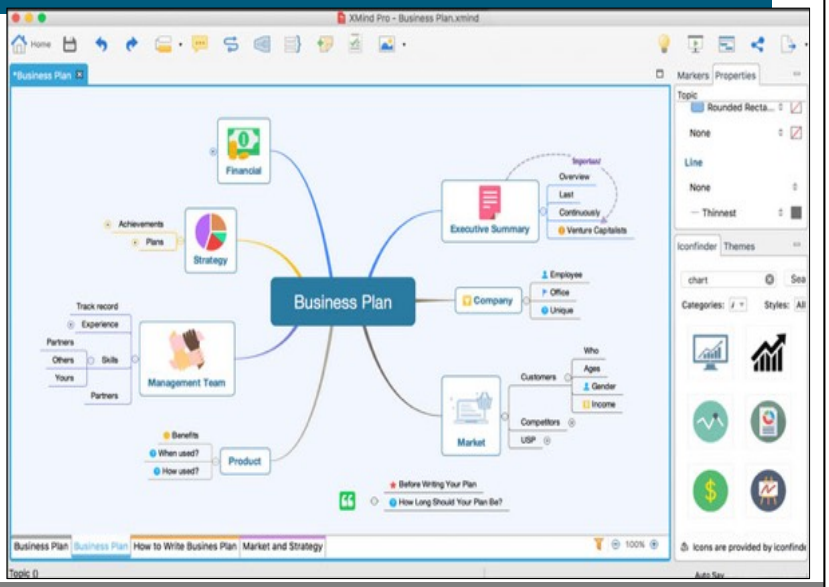
The Gephi visualization tool (https://gephi.org/) is a powerful, free application for data scientists and OSINT analysts with a lot of data to analyze. OSINT researchers such as Nico Dekens (@dutch_osintguy) and Justin Seitz (@jms_dot_py) have demonstrated how to use Gephi to glean relationships in data sets. Above, an image from one of Justin Seitz's blog posts (https://sec487.info/j1) can be seen. While powerful and requiring a powerful computer to perform the hard mapping work, Gephi requires a bit of data science knowledge or a good bit of documentation reading before you can be successful in your analyses.

Image from https://sec487.info/j1, December 11, 2018.

Note-Taking Applications

Examples include:

- Microsoft OneNote
- Evernote
- Google Keep
- MindMaps
  - XMind
  - FreeMind
  - MindDomo



Note-Taking Applications

The note-taking applications provide us the flexibility to record what we feel is important during our investigations. They can be text based or graphical in their implementations and can allow us to embed images and documents in the file to keep all of our notes in a single location. Many of these tools are cross-platform, and some even have cloud-based components. They can also allow multiple analysts to work in the same document at the same time, allowing your team to centralize all of their notes. Below are several of these applications:

- **Microsoft OneNote** – OneNote (https://sec487.info/j3) is Microsoft's note-taking tool. There are versions for Windows and Mac operating systems as well as the cloud-based Office 365 web version. The OneNote documents can be modified by several people simultaneously, so working in groups becomes easier. OneNote documents can contain references to or full copies of documents that you would like to keep in your notes. Analysts can also embed videos and images. Teams that purchase the Microsoft Office software (or Office 365) can leverage other tools in the suite, such as Word and Excel, for additional note-taking and analysis activities.
- **Evernote (https://evernote.com/) and Google Keep (https://sec487.info/j4)** – The popular Evernote and Google Keep applications are widely used for keeping notes. One concern with using them may be their use of cloud storage for data
- **MindMaps** – This class of application is unique, as it is a visual note-taking application that uses hierarchical nodes (i.e., parent notes and child entities) like an outline but usually grouped around a central root node. MindMap software from FreeMind (https://sec487.info/j5) and XMind (https://www.xmind.net/) are cross-platform, Java-based applications that are free. XMind has a professional version for around $100 that adds some additional features, such as additional export/report formats that are helpful. Other vendors have features that may be of interest to you as well (https://www.mindomo.com). These applications allow the embedding of files and images, similar to OneNote. An interesting note about them is that since nodes are hierarchical, you can expand and collapse nodes to display or hide them. This can be helpful because after documenting a significant amount of data in one branch, when you move to the next branch, you can collapse the previous branch and keep your workspace less cluttered.

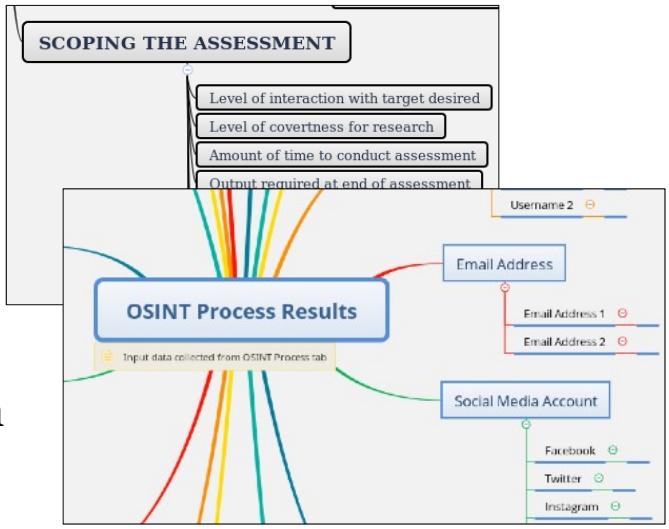
Image from https://www.xmind.net/, August 5, 2016.

OSINT MindMap

There is a MindMap file for you at:

https://sec487.info/id

In it are sample scoping questions and a collection templates for recording gathered data



OSINT MindMap

The MindMap note-taking format may be new to you, and it can be a valuable method for documenting processes, templates, and taking notes during an assessment. There is a MindMap file available for your use at https://sec487.info/id. It has multiple sheets: a process template, a note-taking template, a resources page, and more. Feel free to use these and extend them as your organization sees fit.

Documenting Application

Hunchly Google Chrome Extension

- Created by Justin (@jms_dot_py) Seitz
- Takes pictures of websites as you browse
- Add "selectors" to be highlighted on pages
- Documents URLs and images by page, hash, and time

https://hunch.ly/

SANS Promo code for 15% off "sanssec487"

Documentation Application

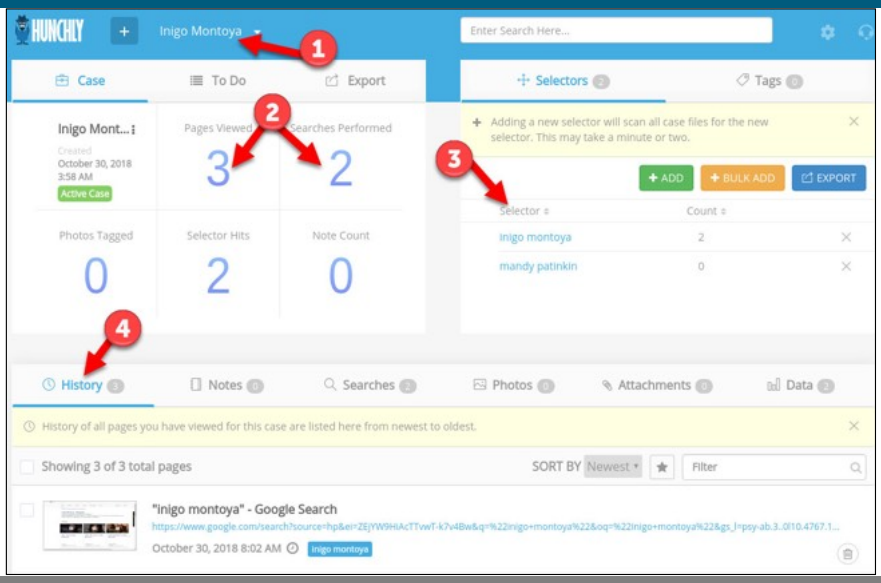
One time-saving tool for analysts to be aware of is Hunchly (https://hunch.ly). Justin (@jms_dot_py) Seitz, an excellent Python programmer and OSINT investigator, created a Google Chrome extension that captures everything an analyst does within the browser. It makes a timeline for when each web page was visited, takes pictures of each web page visited for archival purposes (since web site content can be altered or removed, capturing what an analyst saw or did not see at the time of the assessment can be very important). Justin actively develops this $130 US dollar tool.

Since Hunchly records an offline copy of every web page visited, has the text of each page in a local version and it can search for keywords that are important to your investigation. Hunchly's "selectors" allow analysts to add keywords for Hunchly to search for on each page. It will then show the number of times a certain selector appeared on each page.

Hunchly also creates an automatic timeline for when each of your DuckDuckGo searches or web page visits occurred. It is easily exportable in Microsoft Word format or a compressed Zip file of all the actual, offline copies of the pages and selectors. Oh, and your data, all of it, stays locally on your computer instead of being sent to the internet to some cloud server.

Image from https://hunch.ly/, June 24, 2018.

Hunchly Dashboard



Hunchly Dashboard

When you run Hunchly on your system, you connect your Google Chrome browser with the Hunchly dashboard using a Google Chrome Extension. This software sends content to the dashboard and the Hunchly database for analysis, tagging, and storage. To see the progress of your work, launch the Hunchly dashboard. At the top of the page (arrow 1) you can view and switch "cases," which is how the software logically separates different work. When you begin a new assessment, you create a new case.

The dashboard also displays overviews of what it has collected (arrows marked 2 in the image above). The software can also highlight and tag your work when it finds certain text-based strings if you enter them into the "Selectors" section of the application (arrow 3). These selectors can be entered at any point in your work and, if you choose to enter selectors during your assessment, Hunchly will retroactively search its content and tag existing data with the new selectors you added. You might do this if you discover the name of a new suspect or domain during an investigation. Adding it to Hunchly can highlight content that you may already have collected but didn't know you had.

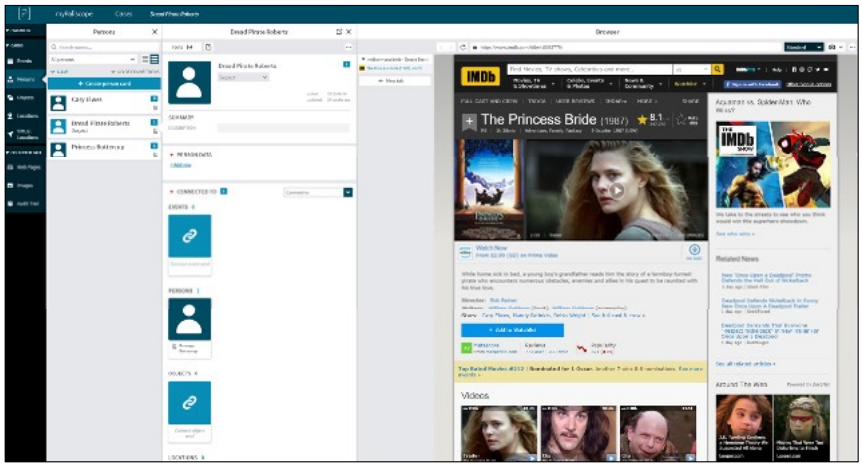
Along the bottom of the image above, you can see briefly how Hunchly notes the history (arrow 4) of what your browser has sent to it, along with notes, tagged photos, and more.

Paliscope

A macOS and Windows case documentation tool

Built-in surface web browser and dark web browser

Make "cards" and link them



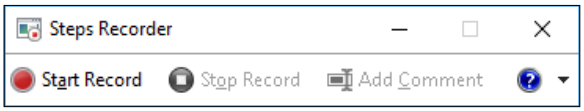
Paliscope

With both a built-in web browser and a built-in dark web browser, the Paliscope (https://www.paliscope.com/) tool can be useful for investigators. Analysts create "cards" with data about people, locations, events, and objects within the applications and then link them together to show relationships. Images that are captured are automatically examined for metadata. Paliscope has a free version for law enforcement and paid versions for other OSINT people.

Documenting Application

Microsoft Steps Recorder

- Free and in modern versions of Windows
- Records:
  - Desktop
  - All keystrokes
  - All mouse movement and clicks
  - Dates and times when each action was performed



- Saves files in MHT format
- Output is a beautiful report of everything done in the system

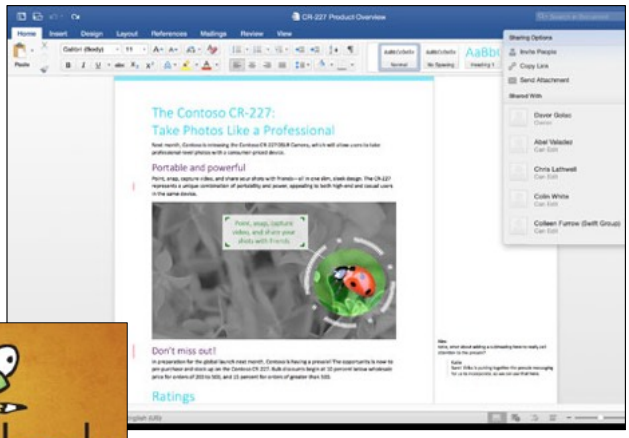
Documentation Application

Microsoft Windows operating systems have a built-in screen-recording tool called Steps Recorder. It records all clicks, keyboard presses, and desktop contents with the date and time when they occurred and then stores all of this in chronological order in an MHT file for later use. The file shows all actions performed on the operating system, when they were done, and an easy-to-read document. Since this is a free option, if you perform OSINT on a Windows host, or in a Windows virtual machine, you might wish to consider using this tool to record what you do.

Word Processors and Text Editors

Examples include:

- Microsoft Word
- OpenOffice Writer
- Notepad++
- Sublime
- Leafpad
- vi/vim



Word Processors and Text Editors

This is a category of applications that most people are already familiar with: word processors. Common applications in this arena are Microsoft Word, Notepad++, Sublime, Leafpad, and vim (not Emacs! ☺). There are hundreds of other possible apps that can be used for manual recording of notes, images, and analysis. These tools range from free to moderately expensive and can be used to create reports for your customers.

Images from https://sec487.info/jf and https://sec487.info/jg, August 5, 2016.

Screenshots

During our assessments we will need to take pictures of content on our computers

There are a variety of built-in and after-install tools that can assist analysts

Example recon-ng screen output

```
[*] 60 rows returned
[recon-ng][test][profiler] > show profiles

  +-------------------------------------------------------------------------
  | rowid | username |           resource           |
  +-------------------------------------------------------------------------
  | 2     | potus    | Bitbucket                    | https://bitbuc
  | 3     | potus    | Black Planet                 | http://www.bla
  | 4     | potus    | Blogspot                     | http://potus.b
  | 5     | potus    | BodyBuilding.com             | http://api.bod
  | 6     | potus    | cafemom                      | http://www.caf
  | 7     | potus    | Break                        | http://www.bre
  | 8     | potus    | CodePlex                     | http://www.cod
  | 9     | potus    | Conferize                    | https://www.co
  | 10    | potus    | cHEEZburger                  | http://profile
  | 11    | potus    | Dailymotion                  | http://www.dai
  | 12    | potus    | COLOURlovers                 | http://www.col
  | 13    | potus    | aNobil                       | http://www.ano
  | 14    | potus    | Etsy                         | https://www.et
```


Screenshots

Since we are discussing recording what we as investigators see, we should touch on taking screenshots. Sometimes during our assessments, we need to capture some tool output in the format it is displayed onscreen. Or perhaps we need to capture just part of a picture, document, or web page to record it. In these cases, we use software to capture a screenshot.

Operating System Built-In Screen Capture Tools

macOS

- **Command-Shift-3** to take a picture of desktop(s)
- **Command-Shift-4** to capture a rectangle of the screen
- **Command-Shift-5** to start recording content on screen

Windows

- Press the "**prt sc**" or "**Print Screen**" button on the keyboard and the desktop will be sent into computer memory and can be pasted into an image editor
- **Alt-Print Screen** captures the active window
- Snipping Tool
- **Windows key-G** to record using gaming bar

Operating System Built-In Screen Capture Tools

Most computer operating systems either have built-in commands to take screenshots (such as the macOS Command-Shift-3 command) or have pre-installed applications such as the Windows Snipping Tool or the Linux KDE kdesnapshot applications. These tools may make an image of the entire desktop (or desktops if you have more than one monitor attached to your computer), or they may allow you to select a region of the screen to capture. In the first instance, where the entire desktop was imaged, a post-processing image tool will need to be used to crop and save the image.

Modern operating systems are also shipping with methods of recording on-screen content. Some of these shortcuts are listed above.

Third-Party Screen Capture Tools

While built-in tools are helpful, third-party tools can be more fully featured

FastStone

Snagit®

Some of the better ones are:

- FastStone Image Viewer
- Snagit®
- Shutter (in your Linux VM)

Image post-processing too!

Third-Party Screen Capture Tools

The screen capture applications that ship with computer operating systems help us with basic recording of screen content. There are, however, some better alternatives that you may wish to consider using.

The first tool is a free, Windows-only app called FastStone Image Viewer (https://sec487.info/jh). It is an amazing resource for not only taking screen captures but for post-processing any image. Did you use another tool to take a screenshot and now you need to blur sections out? Image Viewer can do that. Need to crop, resize, or annotate an image? Image Viewer can do those things and more.

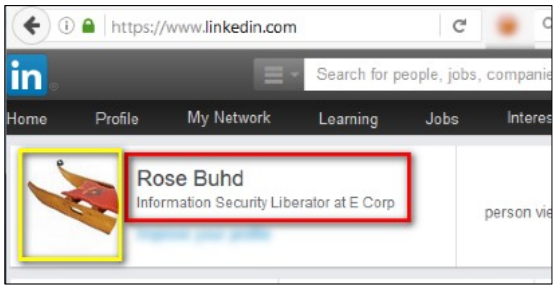
If FastStone's tool doesn't meet your needs, SnagIt® (https://sec487.info/ji) may be able to help out. It has some neat features, such as being able to trim audio and video file content as well as edit images. While it costs around $50 for a license, there is a free trial version that can be used for 15 days.

Your Linux virtual machine has the Shutter application installed. It is a capable tool that can capture regions of the screen and has a post-capture editor too!

Images from https://sec487.info/in and https://sec487.info/jh, September 11, 2016.

Altering Captured Images

- Some customers or jobs require preservation of images
- Sometimes it is important to add content to images
  - Blur sensitive or distracting information
  - Adding circles or rectangles to draw attention to areas
  - Allows you to focus viewer



Altering Captured Images

If you are supporting law enforcement or legal entities, there may be strict guidelines about whether you can modify images. But for the rest of your assessments, you may wish to alter the images that you capture and download. Remember to always keep a copy of the original.

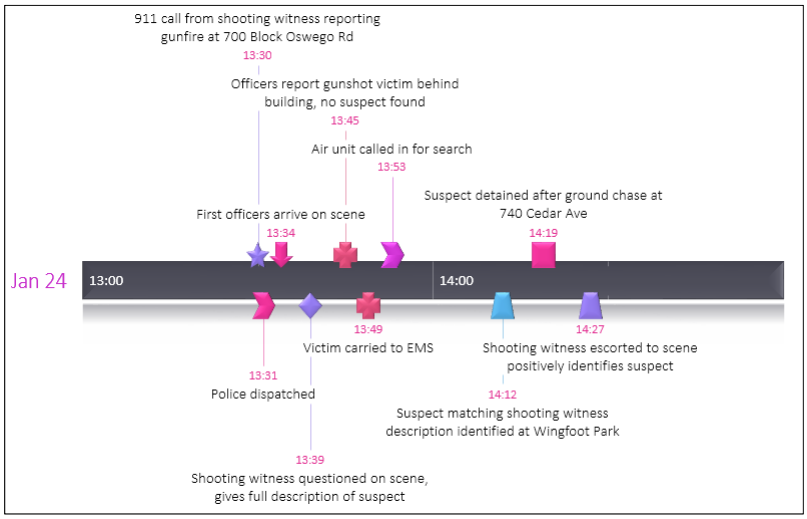
Here are some compelling reasons supporting altering copies of images:

- **Blurring** – Tools like FastStone Image Viewer allow users to drag boxes to blur sections of the picture. We do this to hide sensitive information (for example, passwords in clear text, bystander names and images, and shocking content) and also to decrease the clutter on the page. The first reason is important since you may not know where your report and this image may be viewed. The second issue is that it can help you focus your reader on the information on the page you want them to see and ignore other content captured in your image. In the above slide, look in the upper-right corner and see the section that blurs an add-on.
- **Boxing** – We can use image post-processing tools to put circles or rectangles around certain areas of an image that we want to draw attention to. We can, as shown in the picture above, use different colors for the shapes so that in the text of our report we can refer to them. An example of this might be "As shown in the yellow-boxed area on the left of the image.…" Keep in mind that your report may be printed in black and white or your readers may have color blindness, so you may wish to draw arrows to point at content instead of using boxes.

Timeline Generation

During some investigations we will need to establish what events occurred and in what order -> timelining

Collect, organize and then plot significant dates, times, events, and actions



Timeline Generation

Some of the questions you will be answering for your customers may have a time component. Whether it is "What does my spouse do after work?" or "Piece together the events leading up to the accident," you will need to create timelines. The timelines you make will contain highlighted important events and activities and when they occurred. Adding pictures and additional details to the image can help the reader understand what you are describing.

In case you have not created or seen a timeline before, above is a sample timeline from the https://sec487.info/jk web site. It notes dates and dates alongside important events. If desired, you could add images and links to other data. While we show a timeline from a certain application, remember that other applications can make detailed timelines as well.

Timeline Software and Sites

Use common "office" software such as:

- Spreadsheets in Microsoft Office, OpenOffice, Libre Office
- Slide-making apps such as Microsoft PowerPoint (with add-in)

MindMaps (XMind, FreeMind) have timeline layouts

Specialized timelining applications:

- Aeon Timeline
- Timeline by knight lab

Timeline Software and Sites

There are a variety of methods of making timelines, and each has its benefits. You and your coworkers should choose what works best for your team and start there. There are simple, free templates for Microsoft Office,[1] OpenOffice,[2] and LibreOffice[3] products as well as online web applications[4,5,6] that can generate timelines (careful with OPSEC issues and storing all your case/OSINT data on a third-party web site).

The XMind and other mindmap applications have timeline layouts to organize the nodes in a linear format.

Interested in more links to timeline software? Visit https://sec487.info/ij.

References:

[1] Microsoft PowerPoint (https://sec487.info/dt) and Windows 10 Project Studio (https://sec487.info/du)

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

[3] https://sec487.info/dw

[4] https://sec487.info/jk

[5] https://sec487.info/jm

[6] https://sec487.info/jl

SEC487 Workbook

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

"Mapping Minds and Cases"

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. Goals of OSINT Collection
4. Diving into the Collecting
5. Taking Excellent Notes
6. **Determining Your Threat Profile**
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. Leveraging Search Engines

This page intentionally left blank.

Levels of Openness

"How secretive do I need to be in my OSINT work?"

There are levels of openness:

Level	Description
Overt	Public, non-anonymous
Covert	Low attribution, deniable, secretive
Clandestine	Unattributable and concealed

Levels of Openness

One of the first questions you need to consider when determining your threat profile is "how secretive do I need to be in my OSINT work?" The answer to this question may vary, depending upon the targets of your investigations. Just as police detectives may need to cover their tracks to not tip off their suspect, we too may need to anonymize our network traffic and be alert to various methods of information leakage that could alert our targets that we are looking into their activities.

The intelligence and military communities have three main words to describe their activities: overt, covert and clandestine. We can tie these to our research:

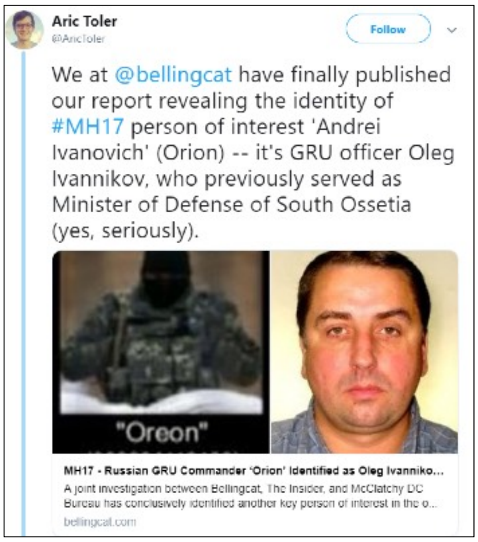
- **Overt** – This essentially means public, non-anonymous activity. Here, we are not trying to be stealthy in our work. It may not matter if the web sites we visit, the tools we use, or even our targets see that we are collecting information.
- **Covert** – A covert assessment, according to https://sec487.info/ie, is "An operation that is so planned and executed as to conceal the identity of or permit plausible denial by the sponsor." Here we expect to use elements of concealment and secrecy to prevent disclosure of our work. Sometimes referred to as "low attribution," we are not trying to be 100% anonymous but are making an effort to make it more difficult for others to know that we are the ones performing OSINT work.
- **Clandestine** - The PDF at https://sec487.info/ie defines clandestine as "An operation sponsored or conducted by governmental departments or agencies in such a way as to assure secrecy or concealment." We call this activity anonymous, where there is no link to the activity and the executor of the work. We usually have many procedures and special systems to hide our activities. If you have ever seen *Mission Impossible* or other spy movie, these are the types of assessments where, in the movies, they say "if you are caught, we will disavow any knowledge of your actions."

The https://sec487.info/36 web site explains the concepts of overt, covert, and clandestine more completely but with a slant toward their use by the United States Department of Defense (DoD).

Publicizing Versus Covertness

Bellingcat and its employee, Aric Toler, went public and revealed the identity of a Russian GRU agent as a person of interest in "*the downing of Malaysian Airlines Flight 17 (MH17).*"[1]

Publicly tying a member of Russia's largest intelligence agency to an event and using your own name is a high-risk event

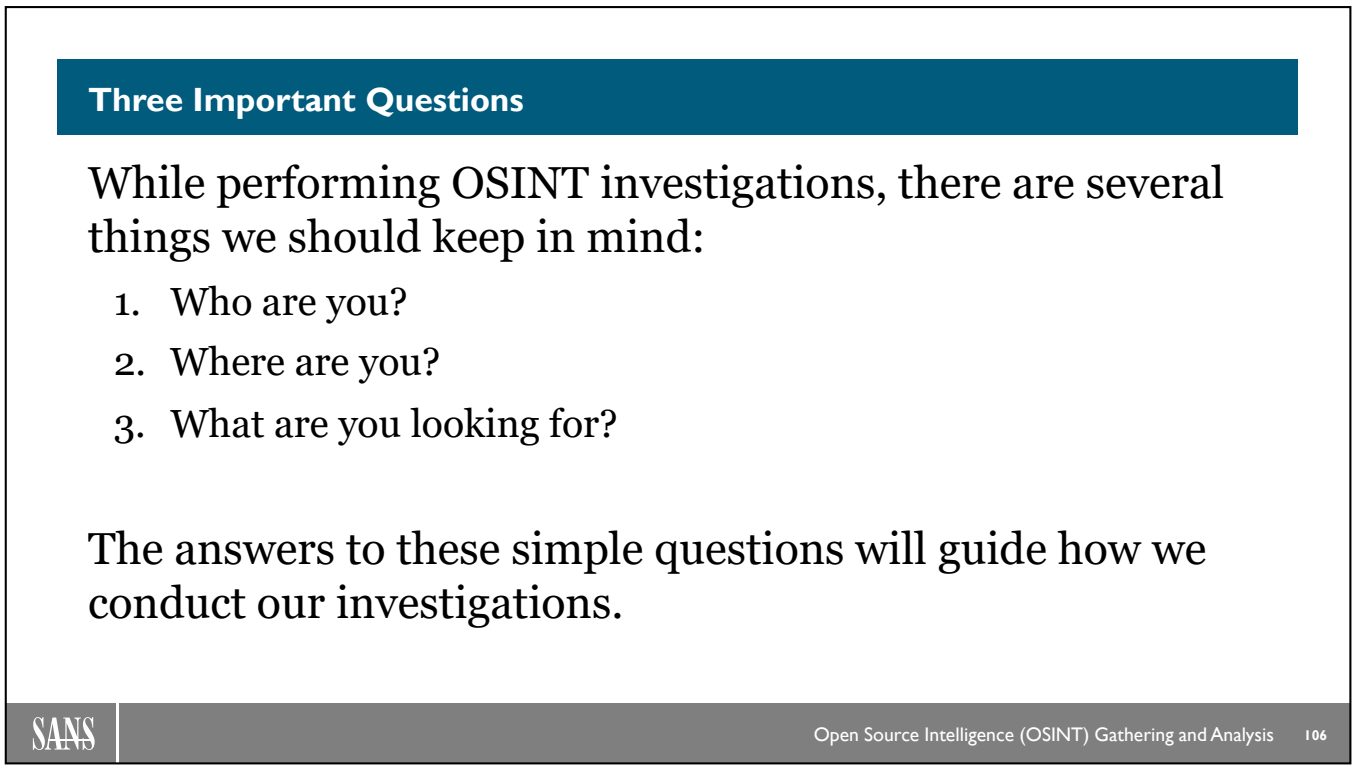


Publicizing Versus Covertness

There may come a time when your team has found something that you want to go public with. It is important to weigh the risk of releasing this information to the public against your safety. Aric Toler and Bellingcat publicly tied Andrei Ivanovich to the downing of Malaysian Airline's flight MH17.

While they may have had the evidence and everything may have been 100% factual, one could argue that they might have wanted to hide the names of those researchers who made the identification so that they and their families could rest a bit easier.

Reference and image from https://sec487.info/jn, December 11, 2018.



Three Important Questions

Now that you understand that there are different levels of openness, you need to figure out what level you need to choose for a given assessment. As mentioned previously, the levels of openness will require you to use different methods of collecting data.

Three questions that can begin to help shape your thought processes are:

1. **Who are you?** Nowadays we have user accounts for everything from our local computers to web sites to services that synchronize our data across different systems. Understanding who you are at each of these locations will help you remain as covert as you wish.
2. **Where are you?** This question refers to your system's network traffic. Where is it coming from? Where is it going? Could it be compromising your OPSEC (Operational Security) and covertness of your assessment?
3. **What are you looking for?** This question is easy to understand. The more specific or restricted data you are trying to collect, the more covert or clandestine you may need to become. An example of a time when this may be a factor might be if you are infiltrating a hacker group's forums to read what they are discussing.

Common Information Disclosure Issues (1)

Where are the places that might leak our data?

Application Issues

- Being logged into a web site, service, or device with a personal or private account instead of an OSINT one
- Cookies, web bugs, and tracking content
- Spurious traffic from browser plug-ins/add-ons
- Unique combination of browser plug-ins and settings
- User agents from browsers and tools

Common Information Disclosure Issues (1)

If our goal is to remain as covert as possible, then we need to understand those places where we could be giving away clues about who we are, where we are, and what we are doing. We will start our analysis at the Application layer.

The Application layer sends and receives data from applications and scripts that we, the users, interact with to get work done. Some places that may leak data here include:

- **Being logged in to a web site or service with your private user account** instead of one specifically used for OSINT can tell others about you, the researcher. A good example of this is how the LinkedIn and Facebook social media web sites tell users who the other users that have been viewing their profile are. Additionally, using a personal account tells the site you are performing OSINT on that you are interested in certain people or behaviors. You may not wish to connect your activities on a web site with your work activities.
- Web sites set **unique cookies** in our browsers and tools. When we replay a cookie that we received when we visited a site for personal use in conjunction with OSINT traffic for work, the web site can tie us to our work, which may not be desired. Web sites commonly track user browsers across the various web sites they visit. Using **web bugs and tracking content**, they can tell if you searched for something in a search engine and then will present you with targeted advertising on various other web sites to entice you to buy their services or products. Tying activities together by ad networks can divulge our OSINT actions.
- The **add-ons and plug-ins that we install inside our web browsers** can be extremely helpful to our work. However, these additions to our browsers can also send traffic about our activities to the author of the add-ons or others and can let them know what we are doing.
- Web sites can sometimes pick out our specific activities on their sites, even if (and sometimes BECAUSE) **we have configured our browsers in a certain fashion**. An example of this would be a web

site that finds a user's browser has privacy plug-ins, user agent changers, proxy switchers, and ad blockers could make an assumption that the user is a security-aware or privacy-aware person. In some cases, where the configuration of settings and add-ons is extremely unique, sites can fingerprint a specific browser using those configurations.

- Each time our scripts, tools, and applications send traffic to web servers, they send information called user agent strings to the destination servers. Inside of these strings is information about the tools we run as well as the versions and the operating systems we use. Check out https://sec487.info/38 for more information.

Electronic Frontier Foundation's Panopticlick

Checks your browser uniqueness using a variety of tests

- Plug-ins
- Screen size/browser settings
- Ad and tracker-blocking
- More!

Electronic Frontier Foundation's Panopticlick

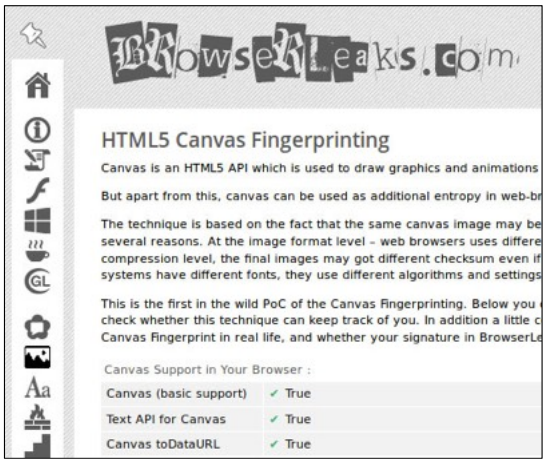
The EFF Panopticlick site, https://sec487.info/jp, can show how unique your web browser is to online sites. It accomplishes this by running a variety of browser tests, from checking if you are using an ad-blocker to how you configured the browser settings to what add-ons are installed. Are you one of those super-secure people that use NoScript, an ad blocker, HTTPSEverywhere, and other security software? Because of those browser extensions and how you configure your browser settings, you may present a unique profile to sites you visit. This combination of your browser parameters can identify your single browser to a site even if you have authenticated to it using different user accounts than normal.

Browser Leakage - browserleaks.com

Free site with a large variety of browser tests using different technologies (Flash, JavaScript, etc.)

Shows what data it found about your browser

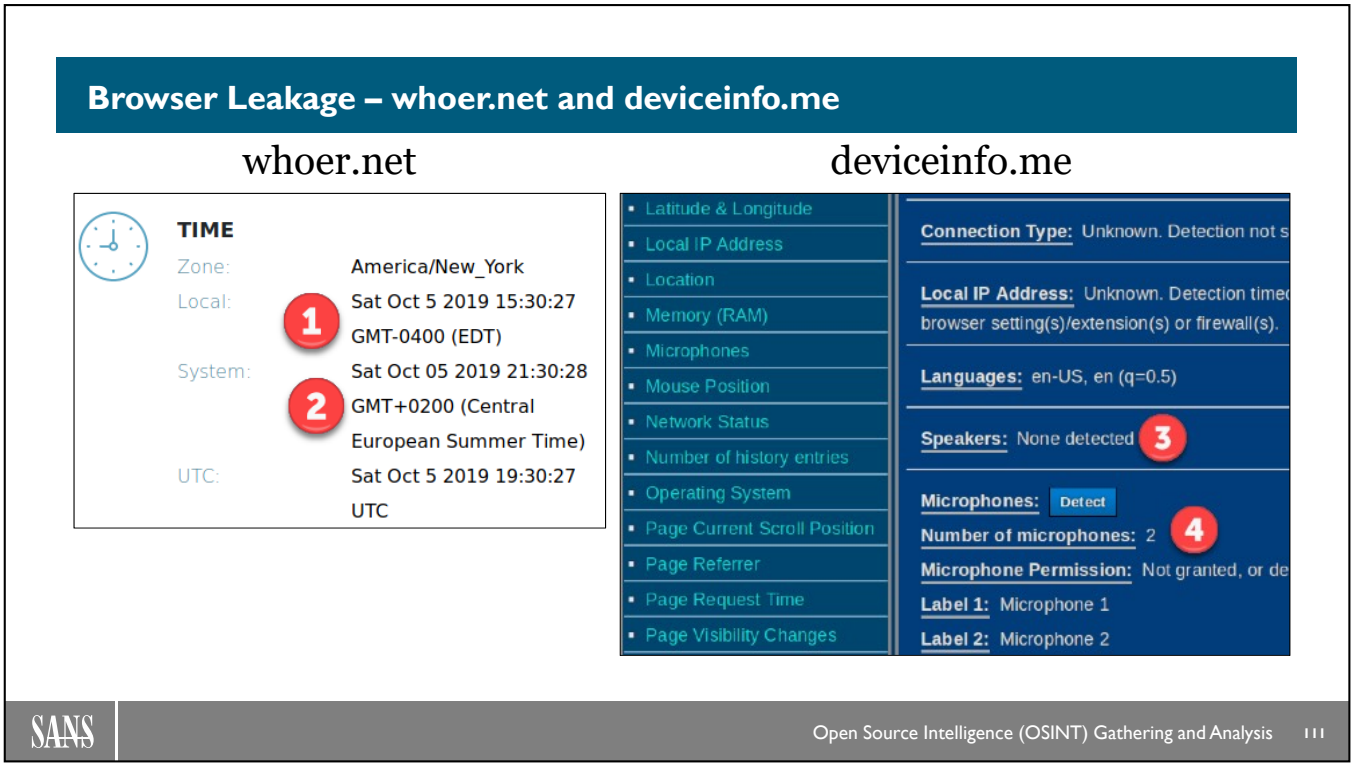
Shows how unique browser is



Browser Leakage - browserleaks.com

The free BrowserLeaks site has a number of tests that you can run a web browser through to see what data it reveals about your system and how unique it is. From JavaScript scripts to Java applets, Flash programs to HTML5 canvas queries, this site has a large number of helpful tests you can use on your applications.

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



Browser Leakage – whoer.net and deviceinfo.me

Another two browser anonymity checkers are https://whoer.net and https://deviceinfo.me sites. Like the others mentioned, these sites display user browser information and will run additional browser tests to reveal places where the browser, its configuration, and its add-ons reveal information.

The image above displays one test it runs on the differences between the local (1) and system (2) times. On the system that the above image was taken on, our host's time zone was set to Paris, France time zone (2) but our host system was hitting the site from an IP address in the Washington DC area (1). The application noted the differences in the time zones, which could give away to a web site that the system may be using a VPN or proxy to surf the internet or that the user is travelling with the device.

The deviceinfo.me site shows other tests of the device, including speaker (3) and microphone (4) discovery. These values, in combination with the others on the page, can reveal a huge amount of data about a system.

Image from https://whoer.net and https://sec487.info/yo, October 5, 2019.

Common Information Disclosure Issues

Network Issues

- IP addresses and network blocks
  - Can give away rough geographic location
  - Also, network providers
- Using Tor
  - Anonymity for you, intelligence agencies, and attackers

Common Information Disclosure Issues
Moving from the Application layer down to the Network and Transport layers, there are several factors to consider when performing OSINT.

- **IP addresses and network blocks** can divulge information about where your system may reside. Internet Protocol (IP) addresses and network blocks on the internet are registered to companies and providers and many times geo-tagged with physical, real-world locations...well, sometimes. Using online services such as whatismyipaddress.com and centralops.net, you can discover country, state, and sometimes latitude and longitude of network blocks. Unfortunately, many times those blocks are further divided, or the information is just wrong. Read the horror story about owners of a Kansas farm who were the victims of erroneous IP address geolocation (https://sec487.info/ii). IP addresses that are registered to companies and agencies also may tie your activities to those organizations. Surf the web for information about a subject from the Interpol headquarters and the destination web sites may make certain assumptions about who you are and what you are doing. Using a Virtual Private Network (VPN) is good protection, but some of those are run by less-reputable companies.
- **Using The Onion Router or Tor** for your OSINT seems like a natural method of anonymizing what location and networks you are coming from. But recent reports have shown that some Tor exit nodes are run by criminals and others by intelligence agencies.[1] Other nodes may be recording,[2] blocking, or altering[3] your traffic.

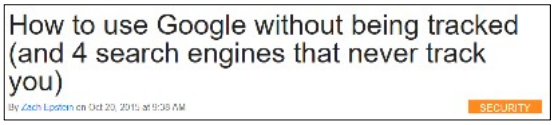
References:
[1] https://sec487.info/3a
[2] https://sec487.info/3b
[3] https://sec487.info/3c

Google Tracking Searches

The services you repeatedly use may be "tailoring" their results for you.

Visit https://duckduckgo.com/privacy

Tailored results may skew your results, analysis, and outcomes.



Google Tracking Searches

Many blog posts and articles have been written about the major search engines collecting data about what is searched for, when, from what IP address, and by whom. Below are some for you to be aware of, and visiting the https://sec487.info/3d page will help you understand how some search engines try not to customize your search results.

Images from https://sec487.info/1c, https://sec487.info/1d, and https://sec487.info/1e, October 2, 2016.

What Can We Do?

1. Understand our system and application network traffic
2. Use proxies and Tor wisely
3. Use appropriate search engines and web sites
4. Clear cache, cookies, and history on our browsers
5. Understand where our applications communicate and what they send
6. Use appropriate OSINT user accounts for work requiring accessing authenticated resources
7. Alter our applications so they provide inaccurate data to sites and network devices
8. Create solid processes and follow them

What Can We Do?

If we understand all the places where our data can be recorded and that information can be spilled to people who should not have it, then what can we do to prevent or at least limit the leakage of information? It turns out there are quite a few things we can do.

Most of our actions are involved with learning how our systems, applications, and add-ons work and what traffic they send to what destinations. When we understand this, we can then choose when to use the appropriate tool and how to send its data to the internet.

Depending on the risk profile of the investigation, you may choose to use certain tools and sources or may decide that they are too risky to use. Create and test your standard operating procedures and processes to ensure that each use case is covered.

OPSEC: Removing Your Data

As investigators, it is helpful to decrease our public online profiles

Open Source project the "Opt Out Doc"

Making it harder for others to use our tools to find us

https://the.osint.ninja/optoutdoc

	A	B	
1	**Public Information Opt-Out Guide** https://the.osint.ninja/optoutdoc Last Modified: 2017-11-26	We would love to see your comments and suggestions. Please email them to optoutdoc@osint.ninja	1. Many websites request conf 2. If you do not find a profile fo the next site.
2	**Website**	**Opt-Out website**	**Notes**
3	addresses.com	http://www.addresses.com/optout.php	1. Use the web site to find you 2. Fill out form on Opt-Out wel under Removal Reasons) 3. Click "Remove Me". Tips: Do not include your midd
4	archives.com	http://www.archives.com/?_act=Optout	1. Use the website to complete 2. If the website states it locate completing the opt-out form. 3. Fill out form on the Opt-Out 4. Click "Submit".
5	backgroundfinder.com	http://search.backgroundfinder.com/optout.html	1. Use the website to complete 2. If the website states it locate completing the opt-out form. 3. Fill out form on the Opt-Out field blank. Put "Please remove 4. Click Submit.

OPSEC: Removing Your Data

It would be uncomfortable and perhaps dangerous if our targets were able to use the same tools we used to find them against us. As we will find in the coming modules, there are many "people search" and "contact search" web sites on the internet. We use them to find others, and they can use them to find us. To address this threat, Micah Hoffman and an anonymous friend have created the Opt Out Doc. It is a bit.ly short URL (https://the.osint.ninja/optoutdoc) that redirects to a Google Sheet.

The document's primary purpose is to provide directions on how to remove your data from sites. It gives step-by-step actions a person can take and relative difficulty levels for how challenging it might be.

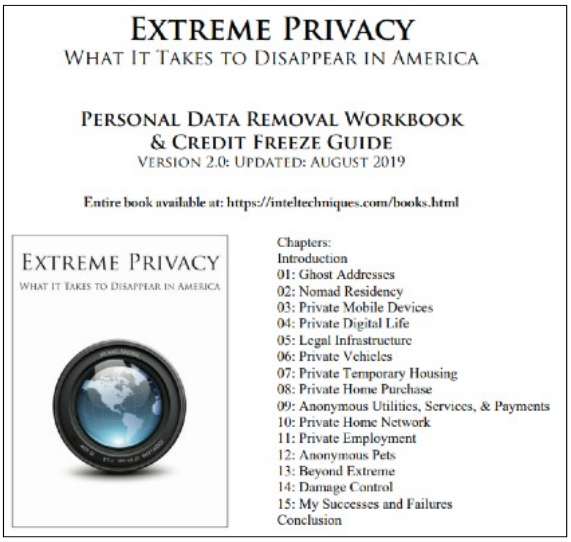
Hiding from the Internet

Michael Bazzell wrote a couple books on increasing your privacy:

- *Hiding From The Internet*
- *Extreme Privacy*

Provides a free PDF workbook

Great resources for protecting yourself personally and professionally



Hiding from the Internet

Michael Bazzell wrote a book (found at https://sec487.info/i0) on removing your data from the internet. This primer is focused on reducing information that can be found about a person in the United States via the internet and other OSINT sources. The book has some simple, easy-to-implement tips and some more-complicated and involved suggestions for the reader. It is an excellent read from an experienced OSINT professional and privacy advocate.

A more recent book from Bazzell, *Extreme Privacy* (https://sec487.info/rn), looks at how to disappear in the United States using extreme privacy measures. This book is for the person who needs to start their life over due to death threats or other reasons. This book explains the suggestions the author provides to his clients to have them reduce, remove, and prevent their information from being collected and served via online sources.

There is an accompanying PDF at https://sec487.info/hz that gives a huge number of action items for the person who wants to protect themselves personally and professionally. This information not only can be helpful to analysts looking to decrease their own internet data but can also be helpful to clients looking to do the same.

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

SEC487 Workbook

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

"Hunchly"

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. Goals of OSINT Collection
4. Diving into the Collecting
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. **Setting Up an OSINT Platform**
8. Effective Habits and Process
9. Leveraging Search Engines

This page intentionally left blank.

What Systems Do You Use?

The assessment systems you use will house your tools and data.

You may have more than one depending upon your targets, your perceived threat profile, and other factors.

You have many choices for where and what you use for your work:

- What operating system?
- Cloud-based, virtual machine or read-only media for the system?
- Do you use a mobile device emulator, burner phone or real smartphone?
- Does your team have funding?

What Systems Do You Use?

Now that we better understand how to determine the threat profile for a given project/assessment, we can more effectively choose a platform from which we can conduct our investigation. If you are doing different types of assessments for a variety of customers, it may be that you have several systems that you leverage during your work, with each used for certain purposes. A good example of this is an analyst who uses a virtual machine–based OSINT platform for the main portion of their reconnaissance, but they may need to tunnel that traffic to a cloud-based system to provide an additional layer of anonymization or a different perspective.

When selecting the systems that we will use, we will need to examine not only the threat profile but other factors, including:

- What operating systems will be used? Does your team have knowledge of one but not others?
- What platforms do your tools run on?
- Will you select a cloud-based, virtual machine–based or read-only media-based platform to perform your work on?
- Will you leverage mobile devices in your work? If so, will you purchase devices or use emulators?
- What kind of financial backing does your team have for purchasing, renting, and subscribing to different systems?

Basic System Requirements

Affordable to use and maintain

Trustworthy – Free from infection, compromise.

Reliable – It and the tools on it work.

"Clean-able" – We need a method to remove prior data.

Special considerations:

- Multi-user/Testing in a team?
- Remotely accessible?

Basic System Requirements

Whatever system we use, it needs to meet certain requirements. Your organization, company, or customer may have specific system requirements, depending upon the laws, regulations, and policies that govern your work. Below are some general items to consider.

- **Affordable to use and maintain** – There are upfront costs when a new OSINT platform is created. Your team may need to purchase virtualization software or rent a cloud server. Many cloud server providers have calculators that can estimate what your monthly bill might be for using their products. You should also consider who is going to install, configure, and maintain the system. These labor costs can be significant, depending on what system you choose to use.
- **Trustworthy** – Some of the web sites you may visit in an OSINT investigation may try to exploit your web browser or host files with malware. Your OSINT system needs to be malware and ad-ware free. Customers would be furious if the work they were paying your company to perform was being stolen by an attacker who had compromised your OSINT platform.
- **Reliable** – Some OSINT assignments are time-constrained. You may be asked to monitor social media sentiment for a current riot or to find a person who went missing recently. When you need to perform an investigation, you need to know that your OSINT platform is there, waiting for you to do your work.
- **"Clean-able"** – After an assessment is completed, you will most likely need to remove the data, configuration files, and artifacts (such as screenshots) you generated during the previous assessment. Comingling one client's results with another's is not a sign of professional work. Because of this, you will want to ensure that the system you use can be easily readied for the next OSINT investigation. Sometimes this means taking a "snapshot" of the pre-assessment system and then, after the investigation is complete, reverting your system to that previous state. Amazon EC2, Microsoft Azure, VMware and VirtualBox, and other servers/software have snapshot features that you can investigate.
- **Multi-user?** – Are you the only person who will be using the platform for OSINT work or will your team use it, too?
- **Remotely accessible?** – Where will you and your team need to access the platform? From home? From inside a certain office? From anywhere around the world?

Considerations for Your OSINT Platform

Consideration	Comments
Affordable	Platforms may have hidden/recurring costs
Trustworthy	Free of viruses and uncompromised
"Clean-able"	Need to remove previous data and artifacts
Availability	On demand? 24x7?
Multi-user	Are you solely using this system, or is a team?
Accessing	Do you need the system local? Can it be remote?
Graphical	Do you need a desktop or just a terminal?
Ease of Updates	How easy is it to add new software and update?

Considerations for your OSINT Platform

Your optimal OSINT platform will be determined by the requirements you, your team, and your customers generate. Most likely, you will have more than one platform, and you may cycle through different systems, depending upon your customer, your costs, and your OSINT goals. There are some factors that should be considered before setting up your first (or your next) OSINT platform.

- **Affordable to use and maintain** – There are upfront costs when a new OSINT platform is created. Your team may need to purchase virtualization software or rent a cloud server. Many cloud server providers have calculators that can estimate what your monthly bill might be for using their products. You should also consider who is going to install, configure, and maintain the system. These labor costs can be significant, depending on what system you choose to use.
- **Trustworthy** – Some of the web sites you visit in an OSINT investigation may try to exploit your web browser or host files with malware. Your OSINT system needs to be malware and ad-ware free. Customers would be furious if the work they were paying your company to perform was being stolen by an attacker who had compromised your OSINT platform.
- **"Clean-able"** – After an assessment is completed, you will most likely need to remove the data, configuration files, and artifacts (such as screenshots) you generated during the previous assessment. Comingling one client's results with another's is not a sign of professional work. Because of this, you will want to ensure that the system you use can be easily readied for the next OSINT investigation. Sometimes this means taking a "snapshot" of the pre-assessment system and then, after the investigation is complete, reverting your system to that previous state. Amazon EC2, Microsoft Azure, VMware and VirtualBox, and other servers/software have snapshot features that you can investigate.
- **Availability** - Will you/your team need to access this resource once a week or every hour of the day? Some systems are easier to keep running for 24-hours-per-day work.

- **Multi-user** - Is it just going to be you using a system or will you have a team leverage it? Using a local virtual machine might be perfect if you are a lone investigator on a project but suboptimal if you have others working the project too.
- **Accessing** – A local system is simple to access since you are using it on your computer. If your team chooses to use a cloud server for the OSINT platform, how will you access it? Where will you and your team need to access the platform? From home? From inside a certain office? From anywhere around the world? Using Secure Shell (SSH) or Remote Desktop Protocol (RDP)? How will you transfer files?
- **Graphical** - Coupled with the preceding items is the idea of using a graphical user interface (GUI) to interact with your OSINT platform. Some of the tools we use in OSINT require a GUI, and others work very well from a command line. An additional consideration is the technical level of your investigators. Are they comfortable navigating using just an SSH terminal window, or do they need a desktop?
- **Ease of Updates** - This is a big one since you will want to keep your installed software and operating system patched and at the latest feature level. This can be easy for some systems (cloud servers and virtual machines with snapshots) and more challenging with others, like read-only media.

What Kind of OSINT Platform?

Each location for your OSINT system has benefits and limitations.

The main OSINT platforms are:

- Dedicated physical system (new laptop)
- Virtual machines (VMs)
- Cloud platform (VDI, virtual browser, and cloud server)

What Kind of OSINT Platform?

As we consider the requirements for our OSINT platform, one of the biggest is where it will reside and what kind of system it will be. There are a variety of options for solid systems, and we will evaluate the main classes:

- **Dedicated physical system** - In this case, investigators can use dedicated OSINT laptops, mobile devices, and desktop systems to perform their OSINT. Some investigators have been known to purchase brand-new devices at the start of every new engagement.
- **Virtual machines** – Systems using server virtualization technologies such as VMware's ESXi (https://sec487.info/2e) or using laptop/desktop software such as Oracle's VirtualBox (https://sec487.info/2f), VMware's Fusion for Mac OS (https://sec487.info/2h), VMware Workstation for Windows (https://sec487.info/ih), and qemu (https://sec487.info/2i) run virtual machines. Michael Bazzell's Buscador (https://sec487.info/3v) VM is free and preconfigured with many OSINT tools installed.
- **Cloud platforms** - Cloud systems are simply servers that are located in someone else's server rooms. These systems can be preconfigured operating systems with software installed or servers that you need to choose and install the operating system on. There are a variety of costs, methods of accessing, and choices with this type of platform. Some cloud server companies include Amazon's Elastic Computer Cloud, or EC2 (https://sec487.info/2k), Microsoft's Azure (https://sec487.info/2l), and DigitalOcean (https://sec487.info/2m). On these vendors' computers, you can install full server operating systems and VDIs (Virtual Desktop Interfaces), where you use a virtual desktop client operating system or a virtualized web browser.

If you are interested in trying out the DigitalOcean.com cloud servers, you can use the link at https://sec487.info/dz to get a $10 credit.

Buscador - OSINT Virtual Machine

Free Linux, OSINT-focused Virtual Machine

Works with VirtualBox and VMware applications

Preinstalled apps

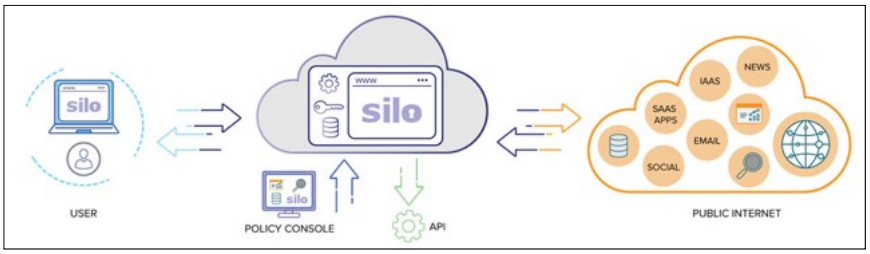
Simplifies tool use through GUIs

Buscador - OSINT Virtual Machine

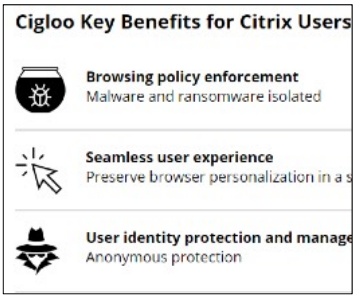
There is a free, OSINT-focused virtual machine that you can download and use for your OSINT research. David Wescott, the creator of Buscador (https://sec487.info/3v), aims to provide an easy-to-use, OSINT research platform that has many of the tools, add-ons, and applications you might want to use pre-installed. You will also find attractive graphical interfaces for many of the command line tools you may use in your OSINT work. Wescott has detailed instructions on the download site on how to install the system.

Virtual Browsers

Use your web browser to reach a virtual web browser in the cloud



Your web browsing comes from the cloud IPs and is virtual/highly remote accessible



Virtual Browsers

Application virtualization has been used in the Information Technology industry for over a decade. Essentially, someone runs a computer program on a system somewhere remote from your location. They send the output of that application to your web browser through special software. You can do almost everything that you normally would in the application if it were on your local system, but it is not. It is installed and running on someone else's computer, most likely in the cloud.

This allows us to browse to web sites and retrieve documents that we may not otherwise have been able to since the web traffic is not coming from our system (it is insulated) and the files would not be retrieved to our computer locally either.

There are a number of web browser isolation vendors in this space, including (pictured above) Authentic8, Web Gap, and Cigloo. We present these three applications as a sample of what is available and are not endorsing the products. For more details on products in this space, read the excellent blog at https://sec487.info/yx.

Images from https://sec487.info/yw, https://sec487.info/yy, and https://sec487.info/yz, October 5, 2019.

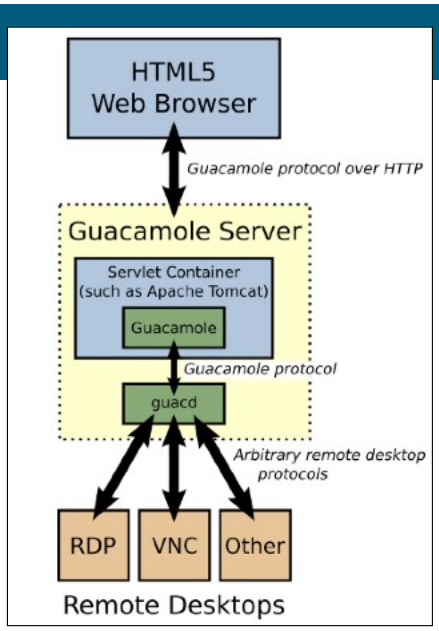
Virtual Desktop Interfaces (VDIs)

If you need a full desktop operating system for your OSINT cases, VDIs may be for you

Run an operating system in the cloud

Your web browser views and interacts with the systems

Players: Amazon, Microsoft, Apache



Virtual Desktop Interfaces (VDIs)

We just learned about companies that run a web browser on a cloud system and allow you to control it from inside your web browser. This virtualization can be taken one step further. Instead of virtualizing just one application, you can virtualize an entire operating system. Using VDIs, you launch a web browser and log in to a site. You choose the system you want to use, and once your system logs in to it, you can interact with a full-blown virtualized desktop from within your web browser.

The Apache Software Foundation has a free tool called Guacamole (https://sec487.info/z0) that is shown above. You can set up a system that runs Guacamole and interacts with client computers you control. This takes a bit of configuration if you are going to do it yourself. If you'd prefer to use someone else's VDIs, Amazon has the WorkSpaces project (https://sec487.info/z1) and Microsoft has Azure Virtual Desktops (https://sec487.info/z2) that you can look into.

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

Additional Considerations

Cost:

- Per user or install?
- Per instance, network traffic, or CPU use charges?
- One-time vs. monthly vs. annual

Cloud Servers:

- Can help to obfuscate your traffic source
- Some sites block cloud server IP addresses

Additional Considerations

There are more factors that should be evaluated prior to choosing OSINT platforms.

Because of the wide variety of pricing models, we do not cover the specific cost of these systems, but price can be a big factor in what you use. Some examples of this variety are software such as VMware may have a per-install license whereas others such as VirtualBox are free. Some cloud servers are a one-time cost but may be unavailable due to network or other issues. Some providers charge per size of the system, per CPU, or per IP address and based upon network traffic. Many also have a "free tier" of server that is free for one year. You now are getting the idea of how challenging comparing pricing models can be.

We also need to think about some additional ways that having a cloud server could be useful. We will mention it briefly here and discuss it more in depth later in the class. First, it can help create a jumping-off point for your assessments that masks your original IP address. If your team performs investigations from their homes, your office, and other places such as hotels, you may want to send all that traffic through a single system to mask the different source locations where your staff is. While that can be a big benefit, some web sites don't allow cloud servers to browse their resources. You will need to generate requirements and choose platforms based upon need.

Choosing the Platform

1. Brainstorm your team's requirements
2. Consult with your customers for requirements
3. Evaluate options
4. Choose the platform(s) that works for how you work
5. Re-evaluate periodically

Choosing the Platform

As mentioned before, you and your team, your management, and your customer will drive what platforms you use during an assessment. Many times, teams will have multiple systems, such as a local virtual machine that can communicate with a cloud-based server. It is recommended that you:

1. Come up with your team's requirements. Do you have restrictions on what can and cannot be installed or purchased? Do you have budget constraints that might help narrow the options?
2. Do your customers have requirements for your setup? They might mandate you store data in some specific method or not use systems in certain places. Talk to them to understand their needs.
3. You will next need to evaluate your options and source solutions.
4. Choose the platform or platforms you think will meet your needs and create/build/buy them.
5. Use the platforms for a while and then re-evaluate whether they still meet your team's needs. Change if necessary.

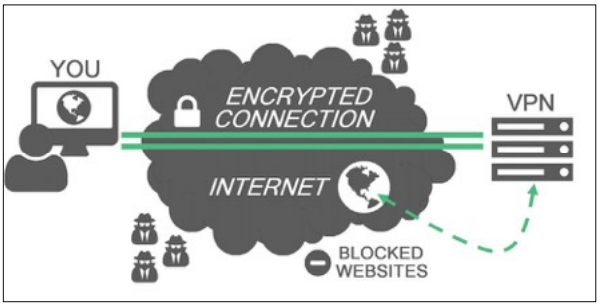
Networking and Virtual Private Networks (VPNs)

Hiding where you are coming from is the main issue here

Can use tethering, a MiFi,[1] or separate internet connection

VPNs

- Encrypt your network data
- Give you a different IP
- Can project presence



Networking and Virtual Private Networks (VPNs)

We will most likely want our home or office network address to be different from our normal, work IP address. We wouldn't want the sites we visited to know that we worked for a government organization or let a competitor know that we were doing business intelligence on them. We can use a separate network for a connection to the internet by tethering our computers to a phone, using a MiFi wireless device, or obtaining a separate network connection through a cable, wireless, or fiber network.

VPNs are useful to our work, as they create encrypted tunnels that our data flows through. This protects our content from being read. Additional benefits of using VPNs include:

- **Different IP addresses** - When you VPN in to a server and send network traffic to it, you effectively browse the internet and interact with sites through that server's IP address. This masks your true IP address. So, if you are working from your home and you don't want people to know that you have a Cox Cable network at your home, you can VPN in to a server and browse the web from it. The VPN provider will know the IP from Cox Cable network but the sites you interact with will not.
- **Different region** - Most VPN providers have servers that can be connected to in numerous countries. This is beneficial to our work, as we can visit sites from an IP address within the country we are investigating.

Drawbacks to using VPNs and other networking appliances can be that they may use well-known IP addresses that may be banned or blocked on the sites you want to visit. Other people on those networks may get those IP addresses banned prior to you using them. Imagine if you are running a tool through a VPN and a site bans the VPN's IP address. Not a problem for you. You just move to a new VPN endpoint, get a new IP, and continue working. But what if I also use that VPN provider, log in, and get assigned your old IP that is already banned at a site I want to visit? I will have to move to a new IP, too.

Images from https://www.amazon.com/Verizon-MiFi-6620L-Jetpack-Wireless/dp/B00NTJKAXG and https://9to5mac.com/guides/vpn/, March 26, 2018.

Reference:

[1] https://en.wikipedia.org/wiki/MiFi

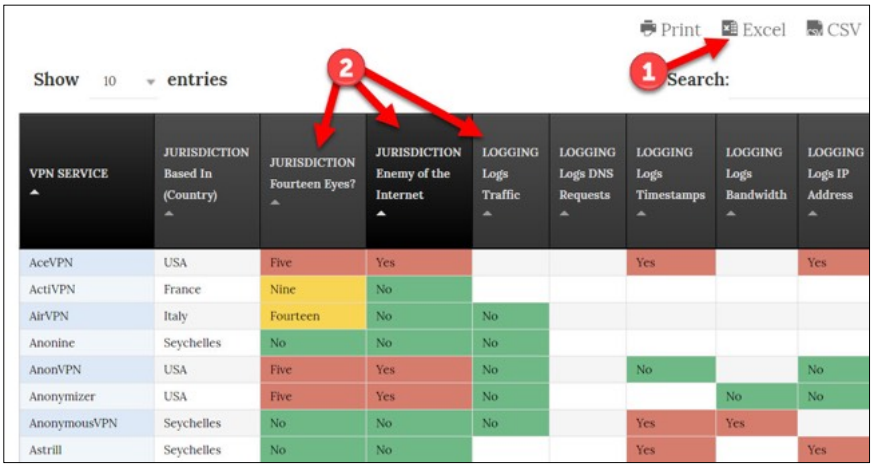
Choosing a VPN - ThatOnePrivacySite.net

There are a huge number of VPN providers

You can examine which may be best for you and your clients

Online and Excel Data

https://sec487.info/go



VPN SERVICE	JURISDICTION Based In (Country)	JURISDICTION Fourteen Eyes?	JURISDICTION Enemy of the Internet	LOGGING Logs Traffic	LOGGING Logs DNS Requests	LOGGING Logs Timestamps	LOGGING Logs Bandwidth	LOGGING Logs IP Address
AceVPN	USA	Five	Yes			Yes		Yes
ActiVPN	France	Nine	No					
AirVPN	Italy	Fourteen	No	No				
Anonine	Seychelles	No	No	No				
AnonVPN	USA	Five	Yes	No		No		No
Anonymizer	USA	Five	Yes	No			No	No
AnonymousVPN	Seychelles	No	No	No		Yes	Yes	
Astrill	Seychelles	No	No			Yes		Yes

Choosing a VPN - ThatOnePrivacySite.net

You may need to get one or more VPNs for your work. The ThatOnePrivacySite.net (https://sec487.info/go) web site has an extremely comprehensive table and Excel sheet (arrow 1 above) for you to use to examine all the options that matter to you and your company. Using their web interface, you can view and filter the different categories and their content to find a VPN or several VPNs that meet your selection criteria (arrow 2).

Create Your Own VPN Endpoint

Use the Community OpenVPN version and a cloud server (Amazon, DigitalOcean, Azure)

Requires you to understand how to configure the system

Blogs and tutorials on how to install these systems can be found online (and in the notes)

May only allow a finite number of IP addresses on some cloud providers; a problem if burned

Create Your Own VPN Endpoint

Your organization may decide that your work is sensitive enough to stand up your own cloud server and run your own OpenVPN server for your OSINT investigative systems to use. This can be cost-effective as, provided you use the community version of OpenVPN and not the Access Server, you only pay for the cloud server resources and not licensing for OpenVPN (https://sec487.info/gr).

A couple of caveats before you decide to install and configure your own OpenVPN server. First, you need to have people who understand Linux and how to install and configure software on it. This can be a challenge when a system is simple, and these OpenVPN servers can get complicated. Additionally, sometimes cloud platforms only allow users a certain number of IP addresses to use on the internet. If the ones that are attached to the server get burned, it can sometimes be challenging to change them.

There are a variety of resources that document how to install and configure these OpenVPN servers in the cloud. A few of them are noted below.

- https://sec487.info/gp
- https://sec487.info/gq

Reference:

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

Image from https://openvpn.net/, March 26, 2018.

Streisand Effect GitHub Project

Sets up a VPN/proxy server in the cloud

Supports: Amazon EC2, DigitalOcean, Google Compute Engine, Linode, and Rackspace

OpenVPN, Tor, OpenSSH, Proxies, IPsec

Streisand GitHub Project

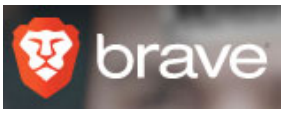
If you would like to set up a cloud-based VPN and proxy server but do not want to have to figure out all the software and configurations to do it, then you might find the Streisand Project (https://sec487.info/hf) helpful. Using the Ansible automation framework, it will configure a full VPN server, complete with a proxy for web traffic, firewall, Tor dark web hidden service, and more! While it requires that you have an account at a cloud provider (e.g., Amazon, Linode, or DigitalOcean), the software for Streisand is free. The only charges you might incur would be from your cloud provider.

To use Streisand, you download it to your local computer, configure it with credentials to alter a cloud environment (usually through API keys), and then run it. When it finishes executing its scripts, you will have a new cloud server VPN to use.

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

Web Browsers

- Web browsers are key tools in OSINT investigations
- Which browsers we use can make a difference
- Some are extensible to allow add-ons that assist our efforts:
  - Increase our privacy and protect our systems
  - Record info
  - Gather additional data



Web Browsers

We use web browsers for much of our OSINT work. Searching, surfing, and collecting web content is made much easier through having a solid, expandable, and secure browser. There are many choices in the browser market, from (mostly) bundled browsers like Safari and Internet Explorer to Google's Chrome (https://sec487.info/2p), Mozilla's Firefox (https://sec487.info/2o), and other specialist browsers like Brave (https://brave.com/).

With modern web browsers working across most of the major operating system platforms, our choices for which to use during our assessments are driven by requirements and functionality.

Our first choices for browsers should be those that are extendable and make our OSINT research easier. Firefox and Chrome are excellent choices that both have extensions or add-ons to increase the things our browsers do during our work. Some of these features increase our privacy and security while we gather OSINT. Others may help us record where we have searched and the pages we have visited. Yet other add-ons may gather more data about the page and its contents or about the sites we visit.

Images from https://sec487.info/2o, https://sec487.info/2p, https://sec487.info/jr, and https://sec487.info/jt, December 11, 2016.

Browser Extensions That Protect

Extension	Purpose
Privacy Badger	Ad and tracker blocker
uBlock Origin	Ad and tracker blocker
Firefox Multi-Account Containers	Isolate cookies and site data to "containers" within your Firefox browser
Location Guard	Allows investigator to spoof location of their browser. Also protects against divulging investigator location.

Browser Extensions That Protect

As OSINT investigators, we will be visiting and retrieving data from web sites that may not be trusted. There will be tracking bugs, malicious JavaScript, and sites that try to track our every query and click. To protect ourselves, we can leverage some add-ons and extensions that will block these attacks and keep our activities more private. We say "more private," as we understand that everything we do on the internet nowadays is monitored and recorded by some devices and/or systems. What we can do is try to prevent these systems from gathering data on our work as much as we can. Let us look at some of the extensions that we can install in our browsers to make our work more secure.

- **Privacy Badger** – Made by the Electronic Freedom Foundation (EFF), this add-on works to block resources loaded from other sites, social media, and advertising trackers. Without a resource/link blocker, you run the risk of loading malicious resources connected to web sites you visit and also allow other sites to track your movements across web applications. More information at https://sec487.info/8g.
- **uBlock Origin** – This fully featured ad blocker is highly customizable and can leverage a variety of lists for domains and web sites that usually serve ads and malware. For those who perform international OSINT, an attractive feature of this extension is that there are ad-blocking lists from several non-US countries. Some people use AdBlock, but they have started allowing ads for those companies that pay them (https://sec487.info/2t), which seems bad for our privacy.
- **Location Guard** (https://sec487.info/2s) – If you are looking to be anonymous on the internet, having websites know what city, state, and country you are sending traffic from is not very stealthy. Location Guard will anonymize your location and allow you to manually set where your browser is located. This can assist in several location-based web sites that OSINT researchers may use.
- **Firefox Multi-Account Containers** (https://sec487.info/hx) – Containerize your browser tabs so that trackers, cookies, and other data cannot pass between them.

There are a large number of other useful add-ons and extensions that can be added into browsers to make them more effective OSINT tools and more secure. It is suggested that you and your team evaluate anything you add to your browsers for security issues and spurious network traffic before they are used in assessments. Some modules have the ability to track, record, and send all user input to the add-on owner. It would be an OPSEC nightmare if your sensitive OSINT investigation data was sent to an adversary.

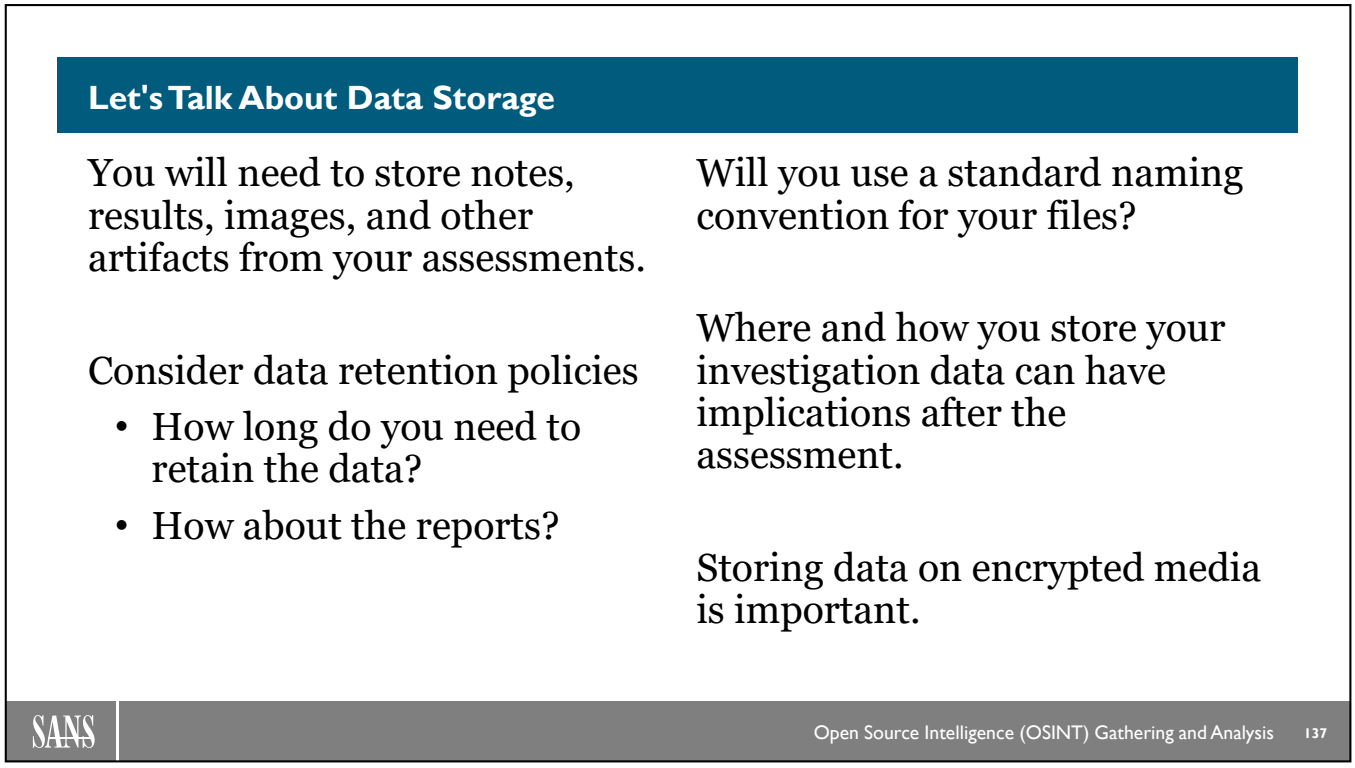
Browser Extensions That Augment

Extension	Purpose
Hunchly	Recording activity
Instant Data Scraper	Extracts table data and saves it as a CSV or XLSX
User Agent Switcher	Alters the browser's user agent string
Download Everything	Media downloader for offline analysis
Search By Image	Sends images to Google's Image Search

Browser Extensions That Augment

As for the browser add-ons and extensions that enhance our investigations, we have a number of interesting options. These are not the only valuable extensions but should serve as places to start. Some of the great extensions in this category are listed below:

- **Hunchly** (https://hunch.ly/) – Has already been mentioned as an excellent data recorder for Google Chrome.
- **Instant Data Scraper** (https://sec487.info/yd) – This extension looks for HTML table data in pages and then extracts it into a CSV or XLSX document, speeding up our ability to collect and store case data.
- **User Agent Switcher** – When we send web traffic to a server, most of the time extra information called the ”user agent" gets sent along with it. This has information about what browser or application you are using, the version, and also the operating system that it is running upon. We can alter this data to try to fool the destination server into thinking we are using a different browser than we actually are.
- **Download Everything** – When performing OSINT investigations, we will want to download images and videos from various web sites. Using an extension like Download Everything (or another downloader) can be a timesaver.
- **Search By Image** – The Google extension for Firefox (you read that right) allows the user to quickly right-click an image in a web page and send it to Google's Image Search.



Let's Talk About Data Storage

As you collect, annotate, and analyze data, you will inevitably need to save that data somewhere. Where you save it will be determined by your system setup and your policies. Some things for you to consider about storage are:

- **Data Retention** – This is, simply put, how long you will store your notes, pictures, videos, and reports. While this will sometimes be determined by your client's requirements (some customers may want you to delete certain information right after delivery of your final report), you and your team should have a default retention policy in place. This will note what types of data your team stores and for what periods of time. Will you keep all data and notes from an assessment or only the final report? Other drivers here may be legal restrictions such as cases that may go to court.
- **File Naming** – Often overlooked as an important consideration, during your assessment you may gather tens to hundreds of pictures, documents and output from a variety of tools. Throwing it all in a single directory with random filenames means that you will need to go back through them later and review each file to determine the type of data and whether it needs to be kept or destroyed. Consider creating a standard naming scheme for the files you collect and perhaps an organizational structure in separate, standard directories.
- **Where Do You Store Data?** – We have already discussed some of the concerns people have about locating data in the cloud, but what about saving data on your local system? On a file share? In a virtual machine (VM)? Where you save content can have big implications after the assessment, especially if your work ends up in legal proceedings or on devices that may be taken outside of your office.
- **Encrypt** – This should not be an option. Each major operating system (Windows, macOS, and Linux) has methods of encrypting hard drives and removable media. Use them to protect your client data when it is at rest.

Storage Encryption

Do you need disk encryption?

- Yes, you do! But all the data I've collected is public
- Aggregation of public data can make it more sensitive

File System Encryption

- Windows – BitLocker
- macOS – FileVault 2
- Linux – LUKS (Linux Unified Key Setup)

Containers – VeraCrypt

Storage Encryption

We already mentioned that using disk encryption is important to prevent unauthorized people from gaining access to your sensitive assessment data. A common sentiment by people is that, since most of the data we collect is public, why should we bother to protect it? There are two reasons: 1) Aggregation of sensitive data makes the information more sensitive; 2) Your analysis and notes will also be stored with this information, and that may have data that should be restricted.

Each major operating system has a mechanism to perform full disk encryption on the underlying disk's file systems.

- **Microsoft Windows** – BitLocker (https://sec487.info/2u) has been built into Windows since the Vista operating system. It is easy to use and well documented.
- **macOS** – FileVault 2 (https://sec487.info/2v) is the software used for macOS since OS X Lion (10.7).
- **Linux** – LUKS, or Linux Unified Key Setup (https://sec487.info/2w), is the main encryption software implemented on most major Linux operating systems (OSs). Most OSs have a simple check box that can be enabled during installation to fully encrypt the hard drive before the OS is installed on it.

There are also container-based encrypted file systems. These are a bit different than encrypting the entire hard drive, as they are encrypted/unencrypted and dismounted/mounted via software installed and run separately from the OS. TrueCrypt (https://sec487.info/ju) was the main name in this category until May 2014, when it was announced that TrueCrypt was not secure (https://sec487.info/2x). A new containerized encryption application, VeraCrypt (https://sec487.info/jv), based on version 7.1a of TrueCrypt, was released. Since it had backward compatibility with old TrueCrypt containers, many users simply moved to that free product.

Using a Mobile Device

Some OSINT techniques work only using a smartphone.

Will your team use a real phone, an emulator, or no device?

Prepaid, pay-as-you-go, Wi-Fi-only, and used phones work well.

Android Emulators

- NoxPlayer
- GenyMotion
- BlueStacks
- Android SDK

iOS (iPhone/iPad)

- Emulators are available

Using a Mobile Device

There are some OSINT techniques that leverage how mobile applications interact with remote servers to discover and connect with users. For these occasions, your team may wish to use an Android or iOS (Apple) device or an emulator that runs mobile operating systems within virtual devices on a computer, or you might choose not to increase the complexity of your platform and not use those techniques.

Inexpensive phones can be obtained from online marketplaces, used electronics sites, and cellular phone carriers. Many carriers also have pay-as-you-go or prepaid options that can be affordable to most organizations and allow investigators to run mobile applications. Each option can be evaluated by your team and management to determine if any and which ones are right for your investigations.

Another option is to use an emulator, which is software that runs on your computer and will act as a mobile device and allow you to run applications just as you would on an actual device. This software emulation can be an efficient and cost-effective approach to mobile device use. There are many options that could be evaluated. Below are some to start your search:

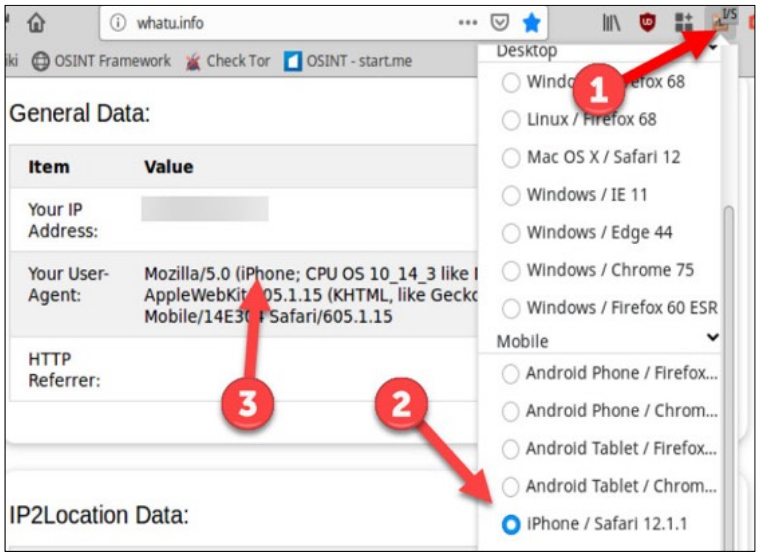
- **Android** – There are several exceptional Android emulators that are free or lost-cost options. These include the Android Software Development Kit (SDK), available for free from https://sec487.info/2z, GenyMotion (https://sec487.info/30), BlueStacks (https://sec487.info/gm), and NoxPlayer (https://www.bignox.com/).
- **iPhone/iPad** – The emulators for iOS devices such as iPhones and iPads range in quality and features but mostly focus on software development. Most analysts agree that using a physical iPad or iPhone device is easiest and fastest for OSINT activities.

In case you need some extra assistance setting up an Android environment for your OSINT practice, Josh Wright, Senior SANS Instructor and course author of SEC575: Mobile Device Security and Ethical Hacking, has an excellent YouTube video on how to set up the Android SDK for mobile testing (https://sec487.info/33).

Change Your Browser User Agent

Using a plugin/addon in our desktop browser (1), we can alter the user agent reported to web sites (2)

This way, we can pretend to be a mobile device (3)



Change Your Browser User Agent

There are many plugins or addons for our desktop browsers that allow us to switch our browser user agent. The user agent is something web sites might use to automatically switch your browser to view a mobile version of their web site instead of a full desktop version. Since this data comes from our browsers, we control it. It is trivial to change our browser user agent so that our traffic appears to be coming from a mobile device.

In the image above, we used the User Agent Switcher addon in the SEC487 Firefox browser to switch the user agent to that of an iPhone. We then visited the site https://whatu.info and viewed what it thought our user agent was, confirming we were using a mobile user agent string.

In some web sites, that is all that is needed to pretend to be coming from a mobile device. In others, this may make your web traffic look suspicious. Turn to the next page to read why.

Make a Desktop Browser Look Like a Mobile One

Change the browser user agent to that of a mobile device

Using the browser Developer Tools, change the device to emulate a mobile device (1), then change the screen resolution (2)

Make a Desktop Browser Look Like a Mobile One

We have already noted several issues from an analyst system that might confuse or be alarming to a web site and trigger an alert that you are doing something suspicious. Changing the user agent string to that of a mobile device might trick simple web sites. More advanced sites will examine the browser user agent string in conjunction with other browser information, like the screen size of your browser.

Our laptops and desktop systems have screens that can be huge in size. A common resolution is 1920 pixels by 1080 pixels. While this is great for watching HD (High Definition) movies, it is a giveaway to web sites that you are not using a mobile device. Most mobile devices have smaller screen sizes. If you do not alter your screen resolution to match the device you are pretending to emulate, then an advanced web site may flag your activity.

To fix this issue, we can use the browser Developer Tools (usually pressing the F12 key on your keyboard will show it). Within the major web browsers, inside the Developer Tools is a place to change the screen resolution of your browser. In the image above, the Firefox Developer Tools were used to bring up the resolution emulator (arrow 1) and then alter it to be the size of a Samsung Galaxy S9/S9+ mobile device (arrow 2). Once this is changed to match the user agent string, our desktop browser looks more similar to a mobile device.

The OSINT Curious Project has a 10-Minute Tip video (https://sec487.info/ye) demonstrating what other things you can do with the web browser developer tools.

Managing Those Passwords

You will have many user names and passwords for your work:

- Sock puppets
- APIs / registered sites
- Paid services

Use a password manager to organize and access them

Managing Those Passwords

As OSINT analysts, we will necessarily have many, many user name and password combinations to track and use. Figuring out a system that works for you and your organization to manage these important pieces of information is, well, important! Too many times people use unencrypted text or spreadsheet documents to track their user names and passwords. Then they find out that those files get stolen, damaged, or backed up to a remote device (without encryption) and their data is no longer safe.

Whether you are managing the login credentials and knowledge-based password reset questions for one of your sock puppet accounts or the credentials for the paid version of some web site that you use during your work, keeping this data safe and accessible is a priority. In recent years, more password-managing options have come to market. Many of them are free or low cost, easy to use, and secure.

Password Manager Options

Local file-based software

- You have control over data and files
- Can share the encrypted file database via Dropbox/Google Drive/Box
- Can be accessed across platforms (mobile, computer)
- Free or bundled with antivirus software

Cloud-based managers

- You manage the data but outsource the security
- Can be accessed across platforms (mobile, computer) via web page
- Usually a monthly/yearly fee

Password Manager Options

The first main choice you need to consider is where you want your passwords stored: on a server in the cloud or in a file you have access to. There are benefits to both types of systems. Below are some of the discussion points to help you make a decision.

- **Cost** - Password managers range from free, to included in other products (such as antivirus suites), to a monthly or yearly fee.
- **Accessibility** - Why use a password manager to make your life simpler if it is hard to get access to the data you need when you need it? You access cloud-based systems through a web browser or mobile application, which makes it easy to send and retrieve content. File-based systems are more challenging to synchronize across computers and mobile devices but can be done using cloud-based storage such as Dropbox, Google Drive, or Box.
- **Software Upkeep** - Cloud-based solutions are simple to "maintain" since you are using their web applications. File-based systems require you to install and maintain programs on your computers. This can sometimes be a barrier to usability.

Getting a Password Manager

Local file-based software

- KeePassXC/KeePassX
  - https://keepassxc.org/ for Windows
  - https://www.keepassx.org/ for Linux/macOS
- PWSafe
  - https://pwsafe.org/

Cloud-based managers

- 1Password
  - https://1password.com/
- LastPass
  - https://www.lastpass.com/
- Dashlane
  - https://www.dashlane.com/

Getting a Password Manager

It doesn't matter which password manager you and your team choose to use. Just use one. It will offload a significant amount of mental energy from you and allow you to make stronger, more complicated passwords. Above are some of the better password managers that you may wish to investigate using if you do not have one already.

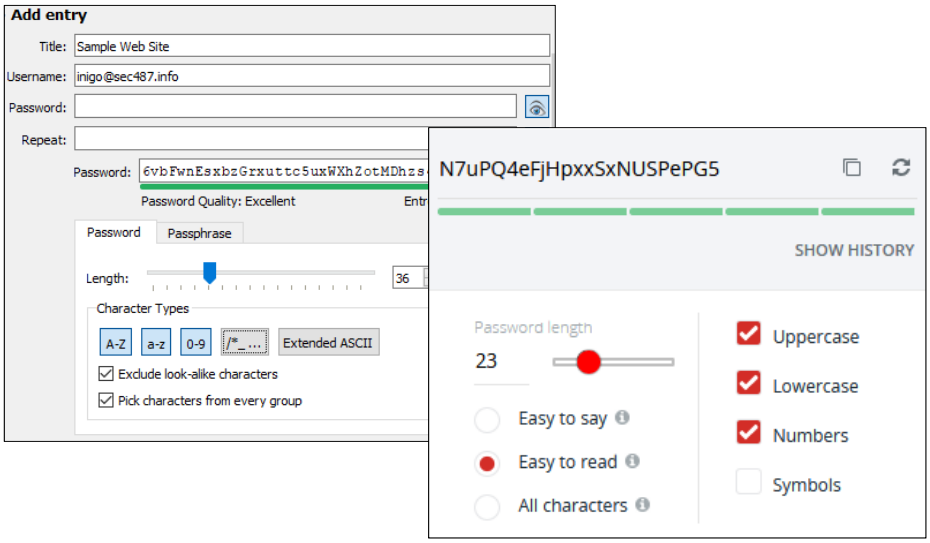
Resources for additional help in choosing the "right" password manager for you:

- Tom's Guide - https://sec487.info/42
- LifeHacker - https://sec487.info/44
- ConsumerReports - https://sec487.info/45

Creating Strong Passwords

Password managers can help us create strong, complex passwords

Use strong passwords and unique knowledge-based questions



Creating Strong Passwords

Our OSINT accounts, be they sock puppets or real, are no different than other user accounts to attackers. They will try to compromise them by guessing passwords or finding out the knowledge-based questions to reset passwords. Help ensure that attackers have a more-challenging time trying to do these activities by ensuring that your passwords are strong, and the knowledge-based answers are random.

Password managers help us with both of these tasks by automatically generating passwords based upon the rules we instruct them to use. They also usually have a notes field that is perfect for entering in the answers for the other information about the account (date of birth, mother's maiden name, knowledge-based answers, etc.).

Images from KeePassXC (left) and LastPass (right) password managers.

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. Goals of OSINT Collection
4. Diving into the Collecting
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. **Effective Habits and Process**
9. Leveraging Search Engines

This page intentionally left blank.

Why Good Research Habits?

When asked to perform OSINT research, there are research habits that will aid the investigator in their work

Failure to address these issues can be factors in:

- Lost work
- Contamination of cases
- Legal and ethical problems

They help us:

- Work efficiently
- Stay in scope
- Stay safe in executing our work
- Prevent ethical conflicts
- Segment customer data
- Maximize our resources

Why Good Research Habits?

You may know a person who just "dives into" their work. Before their customer is finished describing the work to be done, your "friend" may have already started working the project. While this initiative is well intended, it can lead to problems during an OSINT investigation. Our work is about discovering, aggregating, and analyzing pieces of data to form patterns and answer our clients' questions. Using "good" or "healthy" research habits will have many benefits and help us avoid frustrating our customer, contaminating one OSINT assessment's data with that of another assessment, and helping us to avoid legal and ethical problems.

In the coming section, we will discuss how our good habits will aid our investigations.

Getting Connected or Not

Do you connect to your target on social media (or in real life) during an assessment?

- Pro: Can allow you to gain access to large amounts of accurate OSINT information rapidly.
- Con:
  - Alerts target of your work/interest.
  - Can tie a sock puppet account to that engagement

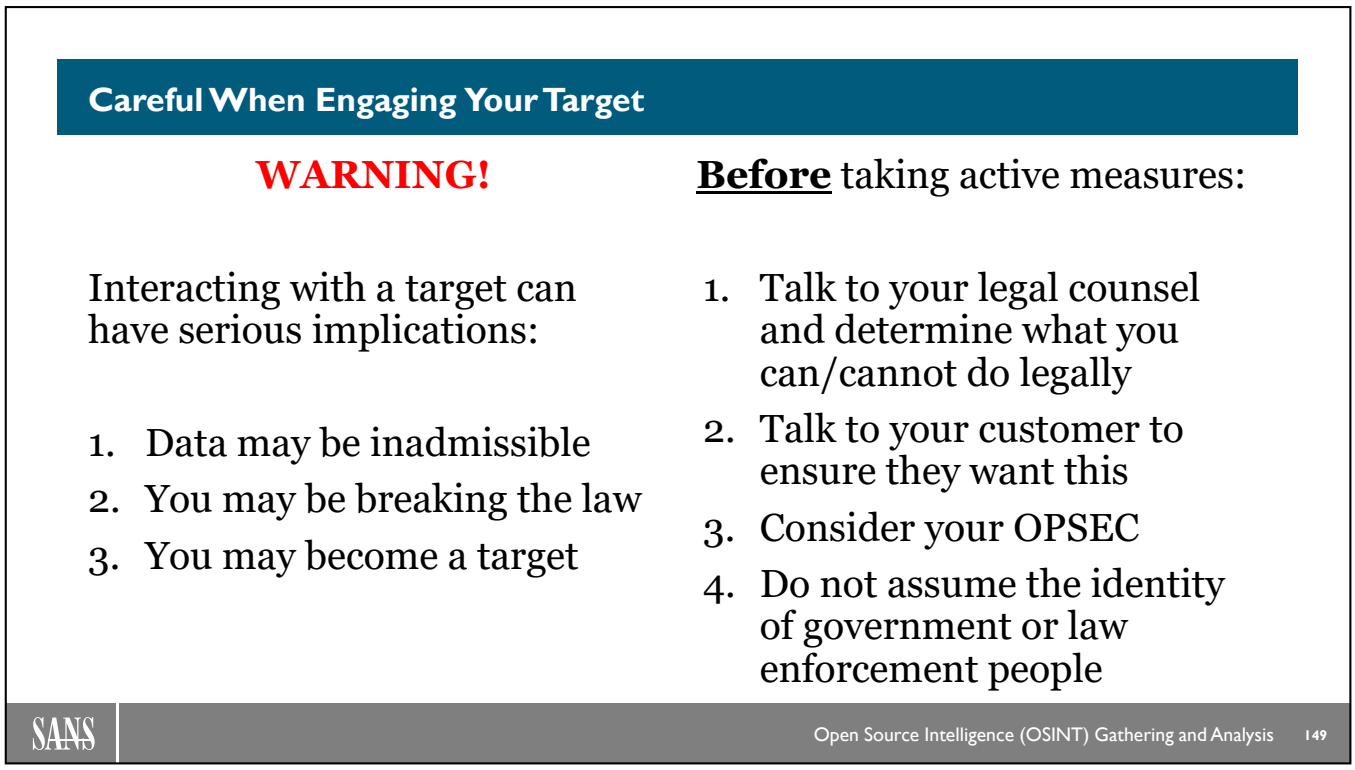
What about connecting to the target's spouse, children, parents, friends, or coworkers?

Getting Connected or Not

The first issue we will cover is whether or not you should connect to your target(s) during your work. This is a topic that your company, your team, and your client should discuss. Within each social network (Facebook, Weibo, VK, etc.) when you connect/friend/link to another's profile, you can retrieve more information than a non-connected user. Getting more information, more quickly can be an efficient method of achieving the goals of the assessment. However, there is also a downside to connecting to your target on social media.

Obviously, by connecting to the subject, you are sharing pieces of your social media profile and account with the target. Hopefully you are using a sock puppet or work profile account for this work and not your personal social media account, so they won't retrieve YOUR friends, colleagues, likes, and such. But recognize that connecting can alert your target that someone is looking for or at them. This could cause the subject to change behavior, become suspicious, or worse (maybe they begin their own OSINT investigation into you!). Connecting to a target may require you to dedicate a sock puppet social media account to that work. If the target knows that that account is not a real, valid person, they could report the account to the social media platform owner or publicly announce it, and you will not be able to use the sock puppet for your work.

Another question to discuss prior to an assessment is, do you connect with friends, family, and coworkers of your target to gain data? This has ethical implications, especially if you are using their family (children, parents, etc.) to gain data about the target.



Careful When Engaging Your Target

Interacting with a target can be a thrilling and powerful method to retrieve data that others cannot. However—and this is a HUGE however—you need to ensure that you are doing it safely and legally so that you, your business, and your family do not get dragged into your work.

Be careful when assuming an identity such as that of a government worker, law enforcement agent, member of the military, or other protected group. The laws regarding this vary across states and countries, and you need to ensure you and your company understand the implications before you take on the persona.

It is STRONGLY ENCOURAGED that you CONSULT LEGAL COUNSEL BEFORE INTERACTING WITH YOUR TARGET.

Finding Sensitive Data

Sensitive data comes in many forms, and you will discover some during your investigations

You need to have a procedure to implement when you find it

When do you stop your assessment and alert authorities and the customer?

Examples of sensitive data:

- Child Sexual Abuse Material (CSAM) – Immediate halt and contact Law Enforcement
- Proprietary or classified data
- Other family (polygamy)
- Alternative lifestyle in country where it is illegal
- Embarrassing photos & videos

Finding Sensitive Data

Perform OSINT work long enough and you will see pictures, watch videos, and read content that you wish you hadn't. People publish so much of their lives on the open internet, and some sites do not protect that data. This makes it easy for OSINT researchers to discover, collect, and analyze sometimes sensitive information. BEFORE you conduct your first assessment, make sure that your team has a process to handle sensitive information collected during an assessment. This process may include using encryption to store the information, notifying the client immediately, and perhaps stopping your investigation and contacting law enforcement.

Some examples of sensitive data are shown in the above slide, and this list is not comprehensive. While many of the topics shown may seem "silly" ("Why should I be concerned if I find out my target is pregnant?"), being ready to handle this data and understanding what your process states you must do is important.

This may go without saying, but we will say it anyway: If you have a government security clearance and find classified data, you have a duty to report it and may need to stop your investigation. Consult with your security officer for proper procedures.

Talk with your team, your management, and your customer about what your process will be when this content is discovered.

Searched Info Used Against Us

Do OSINT while logged in to personal or same sock puppet account?

- You may get targeted ads for the things you search for
- With OSINT searches, it could get "interesting"

Logged in to Google Chrome and perform searches? Google may get the wrong idea about you personally vs. professionally.

- We sometimes visit "interesting" web sites for OSINT

Searched Info Used Against Us

Moving beyond cleaning our systems, we need to ensure that the accounts we use to perform our OSINT are not connected to real-world accounts. We may want to go further and only use certain accounts with specific customers. Social media, search engine, and e-commerce sites are well known for sharing information about users, delivering "relevant" advertising to the user across multiple web sites. We want to avoid this because it shows that some system(s) are tracking what our accounts are doing, and this is poor OPSEC. Beyond advertising issues, as OSINT investigators, we sometimes visit web sites and search for information that we may not visit and search for in our personal lives. Imagine if you performed an assessment using one sock puppet account searching for words like bombs, explosions, and #resist. This sock puppet may be put on watch lists or otherwise tagged as suspicious in certain web sites.

In a similar manner, if you are logged in to a synchronizing service in a browser, browser extension, or other tool and perform searches, those can be recorded under that user account on those servers. Your further search results may be "tailored" based upon your previous searches, and you may miss important data in the future.

Manage Your Time

Most OSINT assessments are performed over a discrete time frame.

To maximize our work:

- Start with trustworthy sites and work toward less trustworthy
- Start with sites that give you the most "good" results
- Start with sites that give you highly relevant data
- Consider increasing scope of investigation as time allows

Manage Your Time

No matter how much time you are given to complete an assessment, it is rarely enough. There are always other sites that should be queried, documents to be analyzed, and avenues of inquiry to be followed. Even if you are given an unlimited amount of time to complete your work, you still need do accomplish that work in an efficient manner.

Start your OSINT investigations by:

- Visiting sites that are trustworthy. If the data that you gather is inaccurate or not trustworthy, then you are using your customer's resources wastefully. Focus on using those sites you know provide users with solid, trustworthy information and move to the less-trustworthy sites as your investigation proceeds.
- Use sites that give you the most "good" results first. Perhaps one search engine regularly shows more results than another. Use the one that allows you to get the most data in the shortest time.
- You should also use sites that yield relevant data. Sure, one search engine may give you more results than others, but if that data is not relevant to your area of inquiry, then time has been wasted.

As you work your way through your preferred sites, move to the less-trustworthy, less-reliable, less-relevant sites and consider increasing your scope, if that is an option, to allow you to move into more obscure areas of a target's data (e.g., moving to do OSINT on the target's family).

Using an OSINT Process

Why should we create an OSINT process?

Can't we just jump into the assessment and "start googling"?

We can, but we will be more effective if we use a process

Your OSINT Process must:

1. Be **reliable**
2. Be **flexible** yet constant
3. Be **repeatable**
4. Maintain the **privacy** of your customers
5. Be **explainable** to others
6. Yield **meaningful results**

Using an OSINT Process

Let's face it: Sometimes when we are told to begin an OSINT task, our minds start thinking of all the different places where we could collect data. We may have the urge to just open a web browser and start ”googling." Jumping in like this can be detrimental to our investigation, especially if we and our work end up in a judicial system and have to explain how we performed our investigation.

Creating and using an OSINT process in your work will provide you the consistency that you need to stay safe online and get your customer excellent results. Your OSINT process must:

1. Be **reliable** in that the process works no matter what you are investigating.
2. Be **flexible** yet constant so that no matter what you are asked to investigate, your process helps frame the actions you will perform. The rigidity comes from you following the process on every assessment.
3. Be **repeatable** so that others on your team or at other organizations can repeat what you have done and obtain similar results.
4. Maintain the **privacy** of your clients by not carrying over data from one assessment into the next. Cleaning your OSINT systems after an investigation is an important step to maintaining this customer and assessment isolation.
5. Be **explainable** to other investigators and to non-OSINT people. Your customer, your team, and perhaps other entities need to understand what tasks you performed in your assessment.
6. Yield **meaningful results** so that your customers get the answers to the questions that they ask you to research.

Walkthrough a Sample Process

To give you a good idea of how a sample OSINT investigation might progress, let's walk through one

Most of the steps in the process have already been discussed, but now we put them all together

This is an example

Walkthrough a Sample Process

Now that we have learned about all the pieces that go into a good process, let us take a minute or two and walk through the steps you might take in an actual OSINT investigation. When you go back to your work, your process may mimic this, have some of these pieces, or be totally different. This is an example to pull all these concepts together.

Ensure a Clean OSINT Platform

Each assessment you perform should be conducted on a "clean" OSINT platform, free of previous investigation data, search results, and artifacts

Possibly create new sock puppets and accounts

Going to court means using new hard drives

Cleaning a system can be:

- Reverting a virtual machine to a known-good state
- Reimaging computers to a known-good state
- Running scripts to remove file, application, and registry artifacts
- Manually archiving and purging data

Ensure a Clean OSINT Platform
A "clean" OSINT platform means that you have removed the artifacts and data from your previous assessments before you begin your next test. If you do not, you run the risk of contaminating new assessment data with old information, making the current assessment confusing, possibly causing you to come to incorrect conclusions and shows sloppy process.

There is no "correct" method of performing this cleaning, as long as you are left with an OSINT platform that is in a fresh state. Examples of different techniques that people use to clean systems include:

- If you are using a virtual machine for your OSINT platform, you can take a snapshot of it prior to beginning your assessment and then revert the system to that pristine state after you have moved all the important data off at the end of the assessment.
- Like reverting a system using a snapshot, you might choose to reimage the computer. This process is usually performed on a host that is not a virtual machine. Prior to starting an assessment, you make an image of the operating system, applications, and all the data on the computer. That gets stored until it needs to be reapplied to the computer after the investigation has completed. Once the image is rewritten to the computer, all the data from the assessment is overwritten.
- Another approach to removing data and restoring the system is to create helper scripts that would be run after an assessment to erase assessment content, flush browser cached data, and restore other parts of the OSINT platform to its pre-assessment state.
- Finally, and most time-consuming, would be to create a checklist of manual tasks that you or your team would need to perform to reset the system.

You may also need to set up new sock puppets. I knew a person who used the same sock puppet accounts for two different clients and it caused the social media site to suggest that his customers may have a friend in common. Not something you want to happen.

Will your client be using your output in court? If so, you may need to load a new hard drive for the work and preserve it after you complete your work.

Acquire Customer Requirements

Once your OSINT system is fresh, you can move to obtaining collection requirements from your customer

The requirements that they provide will shape your actions during the assessment

Consider creating a checklist of questions that you would like answered by your customer

Examples:

- What is your motive?
- Will this support legal action?
- When do you need results?
- What data do you already have?

Acquire Customer Requirements

Collection requirements from your customer drive your assessment. This is the purpose for them engaging you to do OSINT. To ensure that you cover all of the pertinent questions that you need answers to complete a well-executed assessment, take some time with your management and team to figure out common scoping questions that you will ask your customer.

While not exhaustive, consider the stages of your work and think about what you and your team will need to know. These are the questions to ask your customer. Some example questions might be:

- What is the motive for this OSINT assessment? Compliance? Risk determination? Cheating spouse or partner?
- Will the work that you perform be used to support legal action, such as a law suit, or be used in court? If so, your processes may need to be more stringent and more documented than if it will not be used in the legal system.
- "When do you need results?" sounds like a simple question but it is an important one to understand time constraints.
- What data do you already have? Sometimes your customer can give your assessment a head start by providing already known details about the target. Ask for them.

Once this is completed, you may wish to compile the requirements and send them to the customer for acceptance.

Decide on TTPs

Tactics, Techniques, and Procedures used during an assessment may change depending upon the scoping

Create an SOP with the TTPs used in various cases

Map TTPs out in advance to cover different paths

- Need data in 3 hours versus 3 weeks
- Able to engage with the target versus not

Decide on TTPs

Your assessments are like "Choose Your Own Adventure" books (https://sec487.info/jw) in that the overall effort has many paths to get you to the end. You need to choose the appropriate pathway based upon your customer's responses in the scoping conversation. To accomplish your customer's OSINT goals, you will leverage tactics, techniques, and procedures (TTPs) to get you from the beginning to the end.

Capture all the different pathways into a Standard Operation Procedure (SOP) manual that everyone on your team has. It will come in handy if you are summoned to court or have to provide documentation of your process to a third party. Mapping out TTPs in advance also saves time for your analysts, as they know what the next steps are in the process without having to create them after each previous step.

Examples of some possible choices that impact the TTPs you use in your work might be:

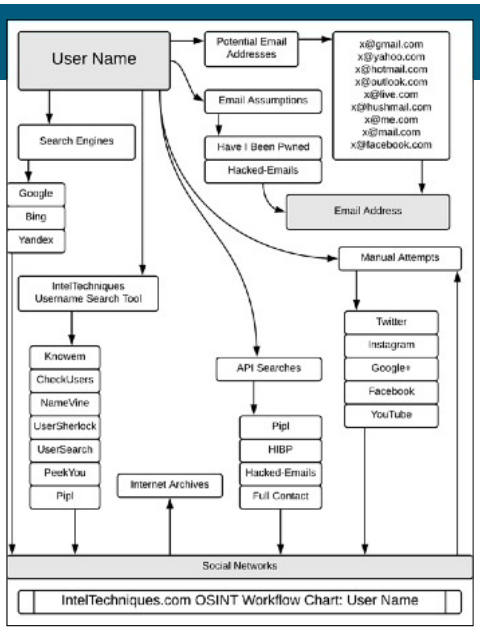
- The time frame when your customer needs the results. Is it hours or weeks? This will help determine the tempo of the assessment and also the depth.
- Are you able to interact with the target? If so, you might be able to “catfish” (https://sec487.info/e4) your target. If not, you may need to use stealthy tactics to assess their activities.

IntelTechniques' Flow Charts Can Help

Michael Bazzell's web site has free flow charts that can be used to augment your processes

Flow charts for research on:

- Email Addresses
- Domains
- Real Names and Usernames
- Phones
- Locations



IntelTechniques' Flow Charts Can Help

Michael Bazzell created and publishes free PDF flow charts that describe how people can obtain data on certain OSINT topics. If you need to find names, phone numbers, domains, emails, or information about locations, his flow charts can be useful. They can be found at https://sec487.info/gs and can be used in your processes or help your team start thinking about how to represent and document your OSINT processes visually.

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

Start Your Note-Taking

New storage area for your investigation data and files (screenshots, downloaded content)

Whatever you are using for a template/SOP/note-taking, create that new version and populate it with customer seed data

"*The only difference between screwing around and science is writing it down.*"[1] Adam Savage

Write everything down/type everything in

Draw pictures and diagrams of connectivity

Start Your Note-Taking

Both you and your company will have your own distinct methods of recording your OSINT notes during your cases. Create a new area in your notebooks, OneNote documents, databases, or whatever it is that you choose to use. Create a new space for any screenshots or files you may download during the case.

Write everything down. Type all the things into your system.

Draw pictures and diagrams of important connections that you see in your work. Relationships between contacts, daily travels on a map, whatever it is for your work...write it down and draw it out in some standard place.

Reference:

[1] https://sec487.info/e5

Gathering Data

This is the core of the assessment

Using your customer's seed data and assessment goals to gather data on your target(s)

Collecting data from all the sources that are relevant and then pivoting on those data points to continue the investigation

Gathering Data

This is the step that most people think of when they think of OSINT. Visiting web sites, performing searches, scraping data, and using OSINT tools. This is the core of what we do. Using the seed data our customer provided, we gather additional related data that we can then pivot on and perform searches using that data. This continues until we have exhausted our time for collection or until we have reached our client's OSINT goals and answered their original questions.

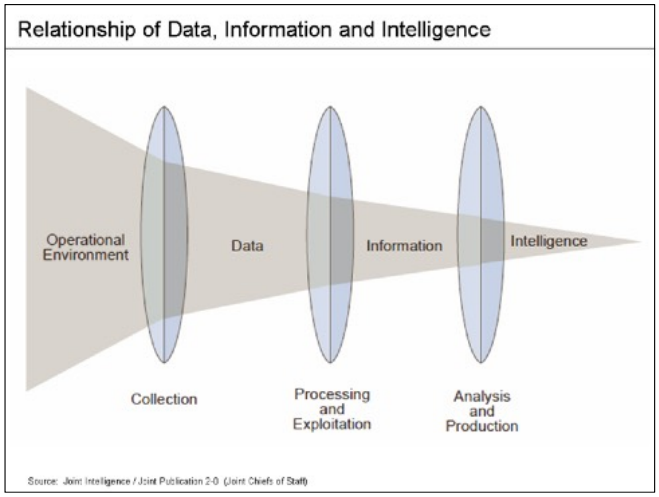
Collect and pivot, collect and pivot, recording your work all the while.

Analyzing Data

Collection can be easy: search and record

Analysis is more challenging:

- What does the data mean?
- Is it related? Is it truthful?
- Avoiding fallacies and bias
- Working with your team



Analyzing Data

Throughout your data collection, you will need to analyze what data you have and transform it into information and intelligence. Just as in the image above, we put data through a series of mental lenses to clarify it: to understand it within the context of our investigations.

In a previous module of the class we spoke about cognitive bias and data analysis challenges. These are important factors to be aware of while we figure out what the data means. We will need to examine the validity and truthfulness of the content and figure out if it should be included, if it should be pivoted on, and if more searching needs to be performed to put context around it.

One of the most powerful methods of analyzing data is to share your analyses with your team (if you have one). Getting others' ideas about alternative ideas and interpretations of the data can be exceedingly helpful in furthering the investigation. Get their feedback and choose the next steps.

Image from https://sec487.info/ib, August 21, 2018.

Repeating the Process

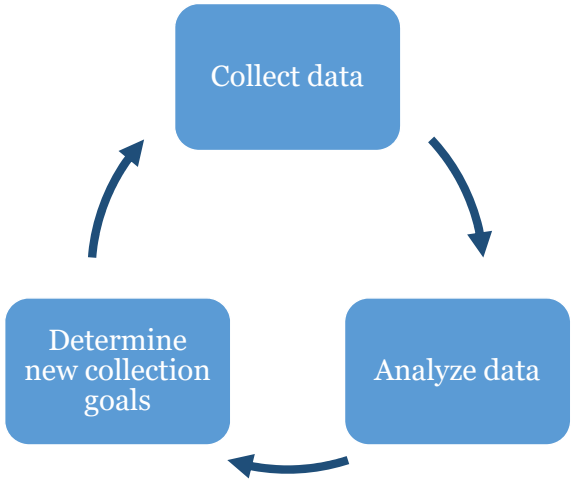
We have a mini cycle inside our larger OSINT process

Gathering data leads to Analysis

Analysis leads to finding gaps

This drives future collection goals

Repeat the process until satisfied



Repeating the Process

The "collection, analysis, new goals, and then more collection" process is cyclical, with each step feeding the next. You will know when you are ready to break out of this cycle when your customer's OSINT goals are satisfied, you run out of time or money, or you have exhausted all avenues of searching.

We have other sections of the course that deal with the specifics of the analysis phase of the process. Here we look for gaps in the data we collected and put it all together, applying reasoning and logic to understand it within the context of the larger assessment.

Create Output for Customer

Sometimes the least-liked portion of the process

Creating meaningful output for the customer is incredibly important

- It is the lasting impression your customer remembers you by
- It helps them move forward with the issue they brought to you

Output can vary from an email or conversation to a report

Create Output for Customer

When your assessment is completed, you need to communicate results and conclusions to your customer. Many technical people loathe this part of the process, but it is important as it is the lasting impression your customer has of you, your company, and their experience with your work. If you provide excellent content, it'll serve as a great reminder to hire you again. A poorly written report may serve to decrease repeat customers and new jobs.

The output could be as simple as an email with your client or may be an elaborate report with embedded screenshots, attachments, and maybe even the hard drive you used to collect the data. Depends upon what you, your company, and your customer want.

Note: Make sure you cover the expected output format during your scoping meeting to ensure that you meet your customer's needs.

Suggest that you and your team create several templates for quick reports and extensive reports and determine what and how you will enter content in them. Having this ready before you need to write the report will speed up the process.

Cleaning Up the System

Delivery of report completed, out-brief briefed

Now to archive or scrub data, revert systems

Prepare for next assessment

Refinement of SOP/Process to improve

Cleaning Up the System

Ah, the time has come when the report has been delivered and your customer has been briefed on your work. Now you look to archive data (if the customer and your work allow it) and revert/clean our systems, readying them for the next assessment.

Another task that is frequently forgotten is to look back over your assessment for places where you can refine your process. Perhaps the scoping could have been tighter, or you need to add new tools to your suite of OSINT tools for the gathering phase. Whatever it is, take a little bit of time to reflect on how the assessment went and if modifications to your SOP and process need to occur.

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

FOUNDATIONS OF OSINT

1. Course Introduction
2. Understanding OSINT
3. Goals of OSINT Collection
4. Diving into the Collecting
5. Taking Excellent Notes
6. Determining Your Threat Profile
7. Setting Up an OSINT Platform
8. Effective Habits and Process
9. **Leveraging Search Engines**

This page intentionally left blank.

Search Engines and OSINT

As OSINT analysts, we need to find data

Since search engines collect and record data, we need to examine their results for connections to our targets

We can direct search engines in what types of data we want

Sometimes challenging to find relevant data

Search Engines and OSINT

If there is one class of web site that OSINT analysts use the most, it would probably be search engines. These systems have automated "crawlers" or "spiders" that visit web pages around the world, categorize them, extract data from them, and allow us to quickly search for our content on those pages.

Even with the vast amount of data being pushed to the internet daily, finding information about our target(s) in search engines can still be challenging. Take, for example, a target named "John Smith" in the United Kingdom. Searching for this common name in Google showed 5,020 hits (https://sec487.info/1n). Sometimes we have challenges when our searches return no data, and sometimes the challenges are when they return way too much data.

Iran Government Just Googled It

In 2009, Iran discovered it could perform a Google search and locate covert channels that the US CIA was using to communicate with informants

Iran used this information to round up and kill informants and then shared this data with its allies

Iran Government Just Googled It

In a public turn of events, two sites (https://sec487.info/ky and https://sec487.info/kx) reported that from 2009 to 2013, Iran discovered the communications network the United States CIA used to communicate to its informants. They found a key characteristic of the web sites that the CIA used and then used Google to find other web sites with those characteristics.

The outcome of this was the killing of at least 30 informants in Iran and Iran sharing these techniques with its allies.

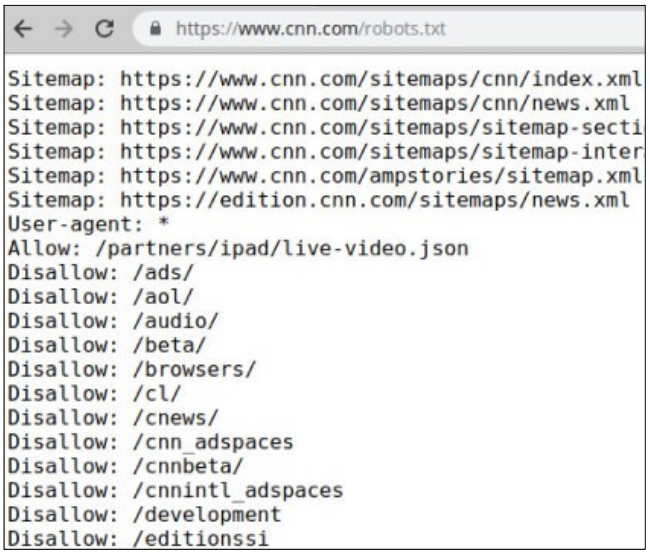
Image from https://sec487.info/kx, December 12, 2018.

Search Engines Do Not Index Everything

Web site owners can prevent search engines from indexing certain content by placing a **robots.txt** file at the root of the domain

Why do we care?

1. Not everything on site is in search engines
2. The places in this file may hold interesting data



Search Engines Do Not Index Everything

Web site owners want search engines to visit their web pages and index their content, catalog their pictures, and ultimately drive users to their web site. However, not all content on web servers should be indexed and easily searchable through a search engine. Some web sites have sensitive pages, such as administrator login forms, or may have content that should not appear on search engine results. For these pages, web site owners can deploy a robots.txt file at the root level of their domain (for example, https://cnn.com/robots.txt, as shown above).

Search engines will send out their web spiders to crawl and index web pages and, when they discover one of these robots.txt files, they will abide by its directions. We see that, in the case of CNN's file, they do not want search engines to index content in the ads, aol, audio, beta, and other directories that have a "Disallow:" in front of them.

For our OSINT purposes, we need to understand that there may be content on the internet that we can retrieve and that may not show up in search engine results because of the permissions in the robots.txt file. Additionally, these files might have sensitive information in them (https://sec487.info/r4) or selectively prevent some content from being indexed (https://sec487.info/r5). As curious people, we may also wish to manually visit these directories that the sites are preventing from being indexed to see what is in them. Do you see any interesting directories in the image above?

There is an OSINT Curious Project video about this at https://sec487.info/r6.

Image from https://www.cnn.com/robots.txt, September 4, 2019.

Which Is the Best Search Engine?

The best one is the one that gives you the data you want

Successful analysts will be agnostic but may have preferred sites they will use before others due to results or features

Experiment with search engines and make your own decisions for the work you do

Which Is the Best Search Engine?

Each major search engine will present the OSINT analyst with results. The "best" one is the engine that gives you more of the data you want and not many false-positive results. Since these search engines may crawl many of the same sites, our results may be similar across the engines. You will need to pick which engines return the best results for your targets, search the most sites that you want searched, and have the best interfaces.

You will most likely use multiple search engines in your work in all cases, as each has its strengths and weaknesses. Get a different perspective on your targets by leveraging search engines in other regions of the world too. Experiment with the various web site search engines to discover what works best for your targets and customers.

Common Advanced Search Operators

Operators	Google	Bing	DuckDuckGo
Negation	- (dash)	- or NOT	- (dash)
Conditional	OR	| or OR	OR
In Title	intitle:	intitle:	t: or intitle:
In Body/Text	intext:	inbody:	b: or inbody:
File Extension	ext: or filetype:	ext: or filetype:	f: or filetype
Domain	site:	site:	site:

Common Advanced Search Operators

The major search engines all have advanced search operators that you can use to refine searches. Google, Bing, and DuckDuckGo use similar operators. We noted them above and describe them in depth below.

- The negation operator can be used to negate a term. An example, `Putin -Kremlin` will show all the results with the word Putin in them and remove the ones that have the word Kremlin.
- The conditional term "`OR`" (or the pipe character (`|`)) can be used to create more complicated searches. As an example, `explosive OR bomb` would show results that have either the word explosive or the word bomb in them. Note that the OR needs to be in uppercase for Google and DuckDuckGo to treat it as an OR. You may ask where the "and" is? You can use it, but most search engines guess that you want to put an "and" between all the terms anyway, so it becomes redundant to use the term.
- The "in title" term tells the search engine to examine the content of the indexed page's title to see if your search terms match. A search for `intitle:"administrator login"` would return pages that had the words administrator and login next to each other in the page's title. The title is marked in the HTML of the page using the <title> and </title> tags.
- Search engines can examine content returned in the body of a web page. A search for `inbody:"Angela Merkel"` would return results for any page that had the words Angela and Merkel next to each other inside the content of the body of the page. The body is marked in the HTML of the page using the <body> and </body> tags.
- Searching for a specific file extension is made easier through the file extension operands. Searching for `filetype:pdf` will return indexed PDF results, whereas other file extensions such as XLS, DOCX, and JPG will return those indexed results.
- The site: operator specifies a certain domain where all the results need to be from. An example of this would be the search `site:sec487.info,` which would return all the results from that domain.

References: https://sec487.info/qx and https://sec487.info/qw

DuckDuckGo.com

Privacy-focused search engine

https://DuckDuckGo.com or https://ddg.gg or Tor hidden service 3g2upl4pq6kufc4m.onion

Doesn't store searches, customize your content, or track your browser

DuckDuckGo.com

We will start our investigation of search engines with one that is dedicated to your privacy and delivering better results, fast. DuckDuckGo (https://duckduckgo.com or https://ddg.gg) has an "about" page that displays the content above. Unlike other search engines that we may use in other portions of our lives, DuckDuckGo focuses on maintaining our privacy and does not track the searches that we perform in order to "customize our search experience." This is both helpful (from a privacy standpoint) and less helpful (as we may get more false positives with basic searches).

While much of the world uses Google as their standard for web searches, we can examine what is retrievable from the internet using an alternate engine. This provides us with contrast to the results from Google and can highlight results that we may not have seen from Google searches.

DuckDuckGo supports and relies on a community of people (https://duck.co) who help to extend the nonstandard/Instant Answer features of its results and also help to translate pages for other, non-English languages. If you are interested in helping improve the search engine for others, you can!

The Instant Answers and !bang responses that DuckDuckGo highlights in its responses provide quick translations of units of measure, currencies, sports scores, and cheat sheets for a variety of applications and operating systems.

Yandex.com

As a Russian corporation, Yandex focuses its content on Asia (Russia) and Europe

Uses many similar search operators as Google, Bing, and DDG

Some unique ones too!

Operators	Use
url: or host:	Results must have certain terms in the URL
domain:	Similar to site:
date: YYYYMMDD	Restrict results to those from a date or range (YYYYMMDD..YYYYMMDD), before (<), or after (>) a date

Yandex.com

While the Russian site Yandex.com has some similar search operators and directives as other engines we have seen, they also have distinct ones that can be useful when searching on their platform. The https://sec487.info/qy web site highlights the terms you can add to your queries that are similar to the other engines: things like using quotation marks (") and pipes (|).

Above, we illustrated several of the more unique terms you can send to Yandex.com to refine your search results. Further details are at https://sec487.info/qz.

Google

Google's search engine is well known, well used, and is recognized as returning relevant results.[1]

It has a wide breadth of systems that it has spidered for content and spiders 20 billion sites every day.[2]

Google indexes web page text and images and allows complex searches to customize result content.

Google

The Google.com web site is well known and is the most-used search engine, with the comScore rankings from March 2016 showing that an estimated 63.8% of desktop computer users chose Google for the search engine needs.[1] Since it retrieves and catalogs data from roughly 20 billion sites each day,[2] it contains one of the widest, most varied databases of content for our searches. That is one of the best reasons for using Google: When we use it to search for content, it usually returns relevant information.

Google's content databases are full of images and text that can be rapidly searched using simple or advanced queries.

References:

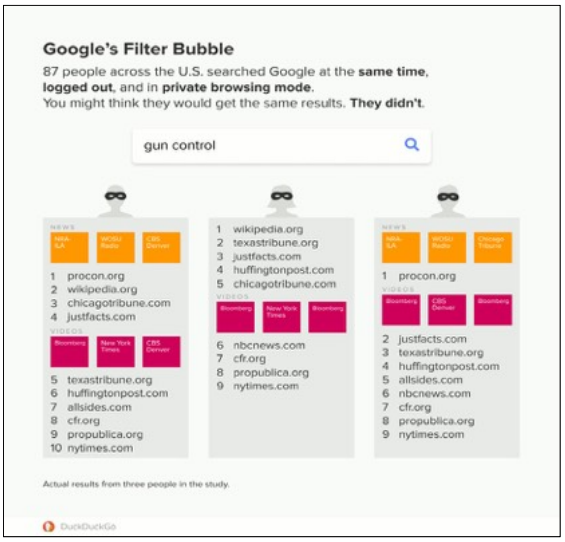
[1] https://sec487.info/en, October 19, 2016.

[2] https://sec487.info/eo, October 19, 2016.

Google's Results Are Mostly Customized

In 2018, DuckDuckGo conducted an experiment where volunteers made the same queries at the same time[1]

1. Most participants saw results unique to them
2. Results within the news and videos info boxes also varied significantly
3. Private browsing mode and being logged out of Google offered very little filter bubble protection



Google's Results Are Mostly Customized

Going back to 2012,[2] studies have shown that Google's search results were customized based on region and user. This "filter bubble" of tailored search results meant that different people searching for the same content would receive varying results. These results were confirmed in 2018 when another study was completed with 87 respondents across the USA. The study "asked volunteers in the U.S. to search for 'gun control', 'immigration', and 'vaccinations' (in that order) at 9pm ET on Sunday, June 24, 2018. Volunteers performed searches first in private browsing mode and logged out of Google, and then again not in private mode (i.e., in 'normal' mode)."[3]

What they found about Google's search engine is shown in the slide above.

Image from https://sec487.info/l3, December 20, 2018

References:

[1] https://sec487.info/l3, December 4, 2018

[2] https://sec487.info/l4, November 4, 2012

[3] https://sec487.info/l3, December 4, 2018

Google Advanced Searches

Operators	Use	Example
before: YYYY after: YYYY-MM-DD	Results indexed before or after specified dates	`Iran missile after:2018-06-29`
* (Asterix)	Wildcard for a word, letter, or phrase	`"former * Anders Fogh Rasmussen"`
word1 around(#) word2	Word 1 must appear within # of words near word 2	`nato around(3) "Anders Fogh Rasmussen"`
#hashtag	Search for records containing a hashtag	`#noextraditiontochina`
#..#	Number or currency range	`2002..2018` `€500..€750`

Google Advanced Searches

Google has a large number of additional search operators that can be used to further refine your searches.[1] Let's examine some and how they can be used in our OSINT.

- **`before: or after: -`** Google has a huge number of records for most of our searches. To filter out ones based upon date, we can use the before and after terms and specify a date in the format YEAR-MO-DAY year-month-day) like 1999-12-22 for December 22, 1999. The dates can be full dates in this format or partials, where you just submit the year and or month.[2] These terms can be used separately (before:2019-1) or together (after:2017 before:2019) to filter your results.
- *** (Asterix)** - This wildcard character fills in for a word, letter, or phrase. So, if you want to retrieve results for all entries that had a reference to the previous positions Anders Fogh Rasmussen had, you could search for "former * Anders Fogh Rasmussen"[3] and you would see references to him being Danish Prime Minister and Sec-Gen of NATO.
- **word1 AROUND(#) word2** - Here we can ask Google to show us results where a certain word or term is within a given number of words from another word or term. So if you wanted to find results with the word "NATO" within three words of the name "Anders Fogh Rasmussen," the query would be NATO around(3) "Anders Fogh Rasmussen".[4]
- **#hashtag** - To search for a specific hashtag that might be used in social media posts, simply type the # character and then follow it with the term you would like to search for.
- **#..#** - This operator allows the user to submit two numbers (and, optionally currency symbols) to search for results within the range of those numbers.

References:

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

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

[3] https://sec487.info/qs

[4] https://sec487.info/qt

Google Dorks and the GHDB

Johnny "ihackstuff" Long started a project to catalog known searches or "dorks" for the community. The Google Hacking Database was born in 2002.

Search modifiers help us locate specific data such as:

```
filetype:xls intitle:password
```


Google Dorks and the GHDB

Since we understand that we can request certain information from Google using the modifiers and operators, we can create a standard set of queries that we may wish to use to search for data about our targets. Back in 2002, Johnny "ihackstuff" Long created an online database of these common searches. He called his project the Google Hacking Database (GHDB).[1]

Johnny wrote the book *Google Hacking for Penetration Testers*, published by Syngress[2] and selling for about $8 on Amazon, detailing the GHDB and how it could be used. He also created a web page so that the community could contribute and vote on Google dorks. His original project is now hosted by Offensive Security within their Exploit Database project (https://sec487.info/1j).

References:

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

[2] https://sec487.info/1i

Using the Dorks

Google dorks are great when you're looking for data

For instance, the WordPress dork:

```
inurl:log -intext:log ext:log inurl:wp-
```

Couple this with a domain of interest by using the site: modifier: `site:example.com`

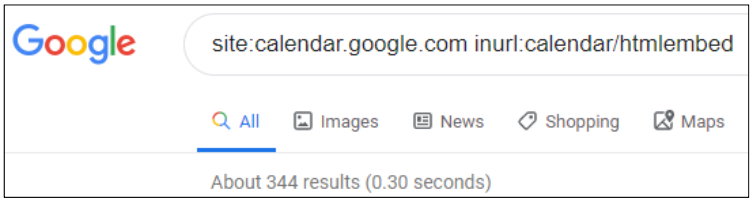
Using the Dorks

Most of the dorks in the GHDB are for finding information from or about systems and domains, not people. So, if your target is a company or you have found a domain or web site that you need to perform some OSINT on, the GHDB can help.

An example query from the GHDB of https://sec487.info/1k is for the following dork: `inurl:log -intext:log ext:log inurl:wp-`. Combine that dork with the domain of your target site using the "site:" modifier and you get `inurl:log -intext:log ext:log inurl:wp- site:example.com`. This will return results for the dork for only the domain you specified.

Creating Your Own Dorks

Find interesting queries.



Did you know that Google calendars can be accessed with:
`https://calendar.google.com/calendar/htmlembed?src=[EMAILADDRESS]`

The Google dork for it might be:
`site:calendar.google.com inurl:calendar/htmlembed`

Creating Your Own Dorks

To make your own Google dorks, either for personal use or to submit to the GHDB, find a URL or other search term you find useful, figure out how to perform the best Google query with it (using the modifiers), and then store it or submit it to GHDB.

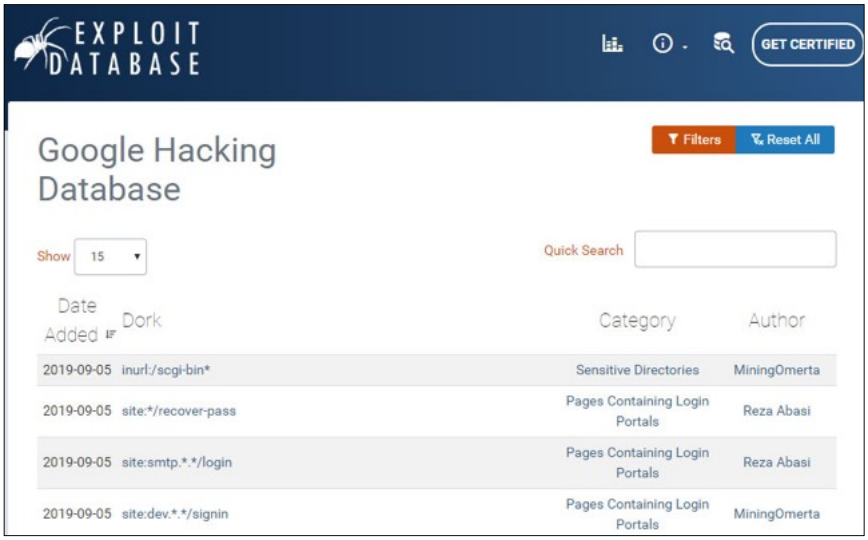
An example here could be that the URL to access Google calendars is `https://calendar.google.com/calendar/htmlembed?src=[EMAILADDRESS]` (where [EMAILADDRESS] is replaced with the target's email address). We can change this into an effective Google dork using the `inurl` modifier: `site:calendar.google.com inurl:calendar/htmlembed`.

Performing this search in September 2019 showed 344 calendars were public (https://sec487.info/ip). Some of these were meant to be public, such as restaurant opening/closing times and sports team's schedules. Others are not supposed to be public and contain entries about how much rent was, when certain people will be at certain places, and much, MUCH more.

Exploit Database GHDB Search

Searching the GHDB is simple

Can also browse the data by category



Exploit Database GHDB Search

Visiting the https://sec487.info/1j URL, people can perform searches to find specific dorks from the thousands that are in the Google Hacking Database. This page also shows the dorks most recently included in the project.

Users can see the search term for the specific dork in the "title" field and also view the category under which the dork was classified. To perform the query, copy and paste the search modifiers (in this picture it could be `inurl:/web/device/login?lang=1)` into a Google search box and submit the form.

Image from https://sec487.info/1j , September 6, 2019.

Automating the Dorks

There are over 4,000 dorks in the GHDB. Running individual queries for each one with your target site is not efficient.

Using tools can get your/your company's IP shunned by Google! Be careful.

There used to be many tools but not now due to API issues.

Automating the Dorks

With over 4,000 dorks in the Google Hacking Database, it is not efficient for a person to manually submit them all against a target of interest. That is where most people turn to automation, and this is what we used to do when Google had a SOAP API that our scripts could communicate with. We could use a number of fast tools to request every GHDB entry, along with our target site. It was beautiful and worked well...until Google closed the SOAP API and made it more challenging to perform automated queries.

If someone with your internet-facing IP address makes automated requests to Google's human search engine (which is against the Terms Of Service per https://sec487.info/1l), it will ask them to solve a CAPTCHA first (CAPTCHAs are some image or audio that the user needs to decode and submit a response to to prove they are human and not a script).[1] If Google thinks you continually perform automated queries against them, your internet-facing IP address can get shunned or blocked from using Google's services for a certain length of time. If your entire company has all their traffic use a single IP address on the internet, "Congrats! You have just blocked everyone in your company from using Google's services!" Be careful and consider using proxies or TOR (The Onion Router) for your automated Google queries.

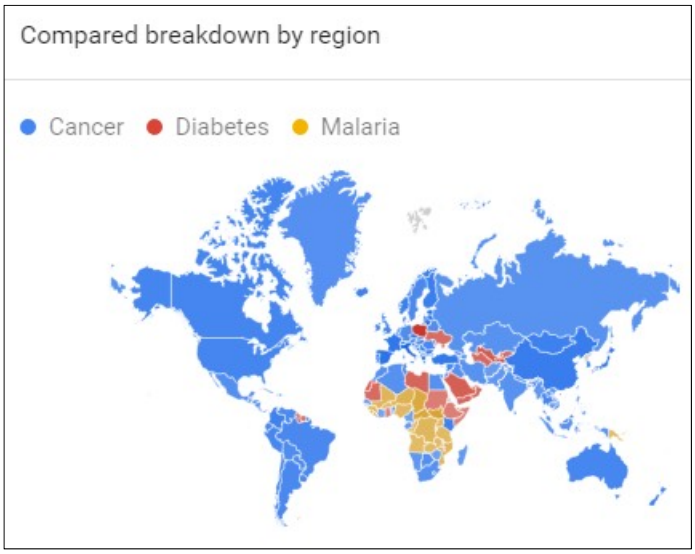
Reference:

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

Comparing Terms in Google Results

Google's Trends tool can be used to compare multiple terms. Determine which term is searched for more, across time and geographic locations.

Can show popularity, brand recognition, topics of interest



Comparing Terms in Google Results

As OSINT researchers, you may be engaged to determine what people in a certain geographic location are researching (googling). Google Trends will help by showing you which terms are more popular, in which regions, and it will even provide sub-searches for each term so that you can better understand what queries people are making using the target terms.

The example above (https://sec487.info/qu) compares searches people around the world have performed using the terms "cancer," "diabetes," and "malaria." The map shows where each is searched for more often.

Image from https://sec487.info/qu, September 3, 2019.

Carrot² Clustering Engine

"Carrot² is an Open Source **Search Results Clustering Engine**. It can automatically organize small collections of documents (search results but not only) into thematic categories."[1]

It searches other engines and categorizes results.

There is a downloadable program if you would like to have a desktop application.

Can help filter results easily.

Blocks access from Tor and cloud servers such as Amazon EC2.

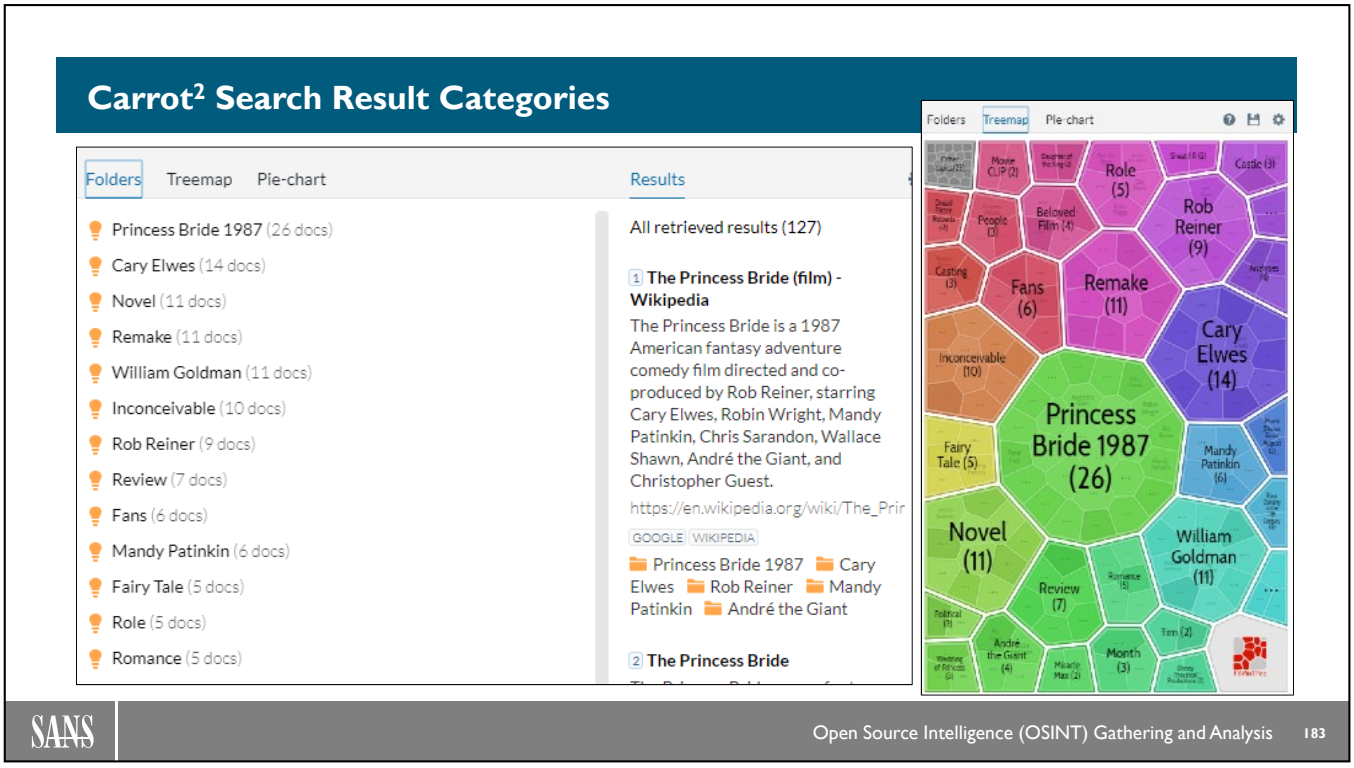
Carrot² Clustering Engine

The Carrot² search engine searches other web search engines and then categorizes the results. It allows the user to quickly filter results using a graphical selection method. Additionally, there is a downloadable client that can be used on macOS, Linux, and Windows systems. The results page of this site (and in the downloadable client) can be organized in a variety of methods, including by which source the data was retrieved from, by date, by category, and by URL.

One issue with using this search engine for OSINT is that it seems that the site blocks traffic to and from Tor networks as well as places such as Amazon EC2 instances. So, if your OSINT system is in the cloud or uses Tor for your investigations, you may have issues using this engine.

Reference:

[1] https://sec487.info/l1, October 30, 2016.



Carrot² Search Result Categories

Filtering the search results is simplified in Carrot² by the frame on the left of the results. Depending upon how you chose to sort the data, the contents of this pane will vary. In the above picture on the left, we chose to cluster the results with "Lingo." More about the algorithms used in the site can be found at https://sec487.info/l2.

The picture on the right in the above slide visualizes the categories from the Lingo clustering as a FoamTree. The more results an entry has, the bigger the foam block. Clicking on the foam blocks will filter the search results accordingly. This can be especially useful to OSINT investigators who are looking for alternative topics to continue their searches.

Images from https://sec487.info/qv, November 12, 2019.

SEC487 Workbook

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

"Searching For IP"

This page intentionally left blank.

SEC487 Workbook

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

"Slacking It"

This page intentionally left blank.

SEC487 Day 1 Summary

Today was about preparation.

We explored documentation, search engines, why we OSINT, what our customers want, and how we collect data.

Tomorrow we will move into collection.

This page intentionally left blank.