

Open Source Intelligence and Cybersecurity

Matthew Danda

Webster University

May 2019

Abstract

Open Source Intelligence (OSINT) is intelligence gathered from publicly available sources such as newspapers, maps, social networking sites, web-based communities, and others. OSINT has become an increasingly relevant source of intelligence for governments, businesses, and criminals alike, as more and more detailed personal information is stored digitally and available online. One significant source of OSINT is social media. When people share information on social networks, they typically intend it for a select audience, as if they were conversing in the real world. However, this information is often available to third parties, and it can be used in unexpected ways. Search engines such as Google provide an easy method for obtaining OSINT. Casual, one-off searches don't necessarily interfere with privacy. However, systematic searches facilitated by automated data collection, selection, and presentation tools can yield significant insight into personal lives. Data analytics can also be applied to perform sophisticated processing and exploitation of open source data. Several academic studies demonstrate how data can be aggregated to provide new information. OSINT tools are numerous, widely varied, and readily accessible. Management tools such as Recon-ng and Maltego are designed to facilitate the process of gathering, organizing, and analyzing OSINT.

Open Source Intelligence and Cybersecurity

Open Source Intelligence (OSINT) is intelligence gathered from publicly available sources such as newspapers, maps, social networking sites, web-based communities, and others. OSINT has become an increasingly relevant source of intelligence for governments, businesses, and criminals alike, as more and more detailed personal information is stored digitally and available online. One factor contributing to the explosion of OSINT in recent times has been the advent of social media and the innate desire for many people to share significant amounts of personal information online.

OSINT is derived from Open Source Information (OSINF), which is the raw data or information. OSINF becomes OSINT when it has been validated as relevant, accurate, and actionable to the consumer. (Williams & Blum, 2018, p8) Thanks to the internet, in recent times OSINF has become widely available and significantly cheaper to obtain. (Eijkman & Weggemans, 2012, p287)

Glassman & Kang (2012, p 674) point out that OSINT is not new; it predates the computer age. Prior to the internet, OSINT was gathered by government-sponsored intelligence agencies from sources such as newspapers, academic papers, and government reports. In modern times, OSINT has become so prevalent that many organizations have been established in both the public and private sector to “study, coordinate or develop new approaches to (the gathering of) OSINF and the acquiring of OSINT.” (Eijkman & Weggemans, 2012, p285)

Best & Cumming (2007) believe that open source information can be classified into one of four categories:

- Widely available data and information. For example, newspapers, magazines, television, and internet-based information. This can also include government reports and official data.
- Targeted commercial data, such as commercial imagery.
- Information from individual experts.
- Gray literature that has limited availability, such as research papers that are only available within certain academic circles. (p6)

Williams & Blum (2018) noted weaknesses in the four categories provided by Best & Cumming, particularly with relation to social media content. (p11) Williams & Blum prefer instead to categorize open source information based on whether it is generated by individuals or by institutions. They further divide these categories into long and short-form information. Long-form information is text-heavy, such as published research or user blogs. Short-form information is typically drawn from posts on social media platforms such as Facebook, Twitter, and LinkedIn. Its value may be limited when viewed individually, however, when reviewed in aggregation, short-form information can reveal patterns and other insights into the subject. (p12)

OSINT is one of five primary disciplines of intelligence as described by Rosenback and Peritz. (2009, p12-13) The other disciplines are:

- SIGINT, or Signals Intelligence, which involves the interception of COMINT (Communications Intelligence) and ELINT (Electronic Intelligence).
- HUMINT, or Human Intelligence, which is gathered from human sources, typically through clandestine operations.

- GEOINT, or Geospatial Intelligence, which is based on the visual representation of activities on Earth.
- MASINT, or Measurement and Signatures Intelligence. MASINT is scientific and highly technical intelligence obtained by analyzing data such as missile plume signatures and uranium particles in the air.

These disciplines are not necessarily unique and distinct from each other, as Williams & Blum pointed out. These discipline can overlap, and their definitions are somewhat driven by the “unique regulatory authorities of intelligence collection agencies than by distinct differences between the collection methods or the material itself.” (2018, p7) Finally, Kris (2017) describes an additional type of intelligence called LOVINT, which is when government officials or security researchers use their intelligence tools to spy on romantic partners.

Glassman & Kang (2012) argue that one of the most important aspects of OSINT is that information can be used in unexpected ways. It is “possible and productive to look for and make connections that are not immediately apparent or are even (initially) counter-intuitive.” (p676) They describe OSINT as being both cognitive and social in nature. It is spread out horizontally across a community, and it is not necessarily maintained by specific individuals and groups. In order to derive intelligence from the raw data, one must make connections between different types of data from various sources. As such, OSINT requires a fluid manner of thinking. (p675)

Three Generations of OSINT

The first generation of OSINT dates back to World War Two and was driven by technological advancement, although not specifically computerized information. (Glassman & Kang, 2012, p675) During that time, intelligence researchers discovered novel ways of gathering

and interconnecting information. Glassman & Kang cite an example in which wartime researchers discovered a correlation between railway efficiency in France and the prices of oranges in Paris, which could then be used to indicate the success or failure of an overnight bombing raid. (p675) The early days of OSINT, which spans from World War Two up until the emergence of the internet, were oriented toward national defense. During that time, significant emphasis was placed on the mere collection of material. (Williams & Blum, 2018, p4)

The second generation of OSINT coincides with the widespread adoption of the internet. As Best and Cumming (2007) so eloquently set forth, “If the 20th century was a century of secrets, the 21st century is the century of global information.” (p2) Williams & Blum (2018) recommend the year 2005 as the beginning of the next generation, as “the bulk of online content shifted to dynamic web pages, user-generated content, and social media.” (p2) That year the Intelligence Community created the Open Source Center, recognizing the need to more effectively capture open source information and train analysts to make better use of it.

The second generation includes social networks and their unintended consequences. When people share information on social networks, they typically intend it for a select audience, as if they were conversing in the real world. As Eijman & Weggemans (p292) point out, this information is often fully transparent to others who can view and reproduce posts. Such openly available information opens up the possibilities for making new types of searches and considering multiple alternative solutions simultaneously. (Glassman & Kang, 2012, p674) Data analytics can also be applied to perform sophisticated processing and exploitation of open source data. (Williams & Blum, 2018, p16)

The third generation of OSINT, as described by Williams & Blum (2018, p39-41), coincides with the evolution to Web 3.0, or the “Semantic Web.” The Semantic Web

incorporates machine learning and automated reasoning to make sense of the ever-increasing amounts of data available for collection and analysis. However, the intelligence community will also face some impediments. Williams & Blum warn that “encryption will become a more prevalent characteristic of third-generation OSINT, as encryption software becomes increasingly pervasive, accessible, and robust.” (p41)

Current Issues Surrounding OSINT

Solove & Schwartz (2015, p11) rightly point out that, “Before the advent of electronic communication, people could easily avoid eavesdroppers by ensuring that nobody else was around during their conversations.” This is no longer the case. Technologies that enable us to record and transmit communications also foster new and sophisticated methods for eavesdropping. In most cases, the subject has no idea that their online behavior has been monitored. (Eijman & Weggemans, 2012, p295) When people upload something to the Internet, they don’t necessarily expect the entire world to actually view it (or have access to it). They have a “reasonable expectation of privacy,” which is a set of principles based on our mutual life experience in the physical world. However, this expectation does not apply to the online world.

The Fourth Amendment was written to protect a man’s house, person, papers, and effects from governmental search and seizure. Unfortunately, it does not address electronic devices, which did not exist at the time it was drafted. Goodman (2016, p94-95) states that the Fourth Amendment does not protect individual’s online privacy with the same rigor as his physical privacy. In fact, Goodman goes so far as to state that any data you post online in any format is not considered private, including data collected by third parties with whom you have an agreed-upon business relationship.

Electronic devices create unique issues related to the temporal and spatial senses. Temporally, humans live and act in real-time. However, the internet can store communications indefinitely and thereby look back in time. From a spatial perspective, humans occupy a single place at a single time. Internet surveillance, on the other hand, can occur in multiple places simultaneously, and the results can be made available in real-time—a feat that is physically impossible in the physical realm with a finite number of human observers. (Koop, 2013, p665)

Search engines such as Google provide easy method for obtaining OSINT. Casual, one-off searches doesn't necessarily interfere with privacy. This is not necessarily the case when using automated search tools. Systematic searches facilitated by automated data collection, selection, and presentation of OSINT may yield significant insight into personal lives. (Koops, 2013, p656) For example, according to Schneier (2015, p40), "Facebook can predict race, personality, sexual orientation, political ideology, relationship status, and drug use on the basis of Like clicks alone."

Several academic studies demonstrate how data can be aggregated to provide new information. Kanakaris, Tzovelekis, & Bandekas (2017) built a system that used publicly available data from Twitter and Instagram to successfully track an unsuspecting person's location and predict their future locations. They were able to do this by combining real-time data from multiple sources (namely, Twitter and Instagram) and analyzing the results.

Along those same lines, Pellet, Shiaeles, & Stavrou (2018) used OSINT from Twitter, Facebook, and Instagram to successfully profile the movement of a person. They gathered information from posts made by the target and by the target's friends, and then they applied machine learning technology to predict the target's current and future location. The authors claim a 77.72% estimated accuracy, even without using GPS data in their model.

Bugs in software may also enable attackers to uncover personal information. Khanna, Zavorsky, and Lindskog (2016, p460) described an incident from Facebook in which their “Download your Information” application unwittingly enabled sharing of email addresses and phone numbers of over 6 million users. Given the vast amounts of data collected and aggregated in today’s society, Schneier (2015) argues at length that it is essentially impossible to remain anonymous.

Government agencies can leverage OSINT to improve the safety and security of the population. However, accountability must be enforced. Eijman & Weggemans (2012, p296) readily admit that state accountability mechanisms have struggled to adapt to the online open source culture. Just because data is accessible doesn’t mean that it is ethical to use. (p291) “From a human rights perspective, the gathering of OSINT demands proper checks and balances.” (p286)

Using OSINT

This section describes how various organizations use OSINT to further their goals. The intelligence community gathers OSINT for purposes of national security. Williams & Blum (2018, p13-14) describe the four steps of the OSINT operations cycle: collection, processing, exploitation, and production.

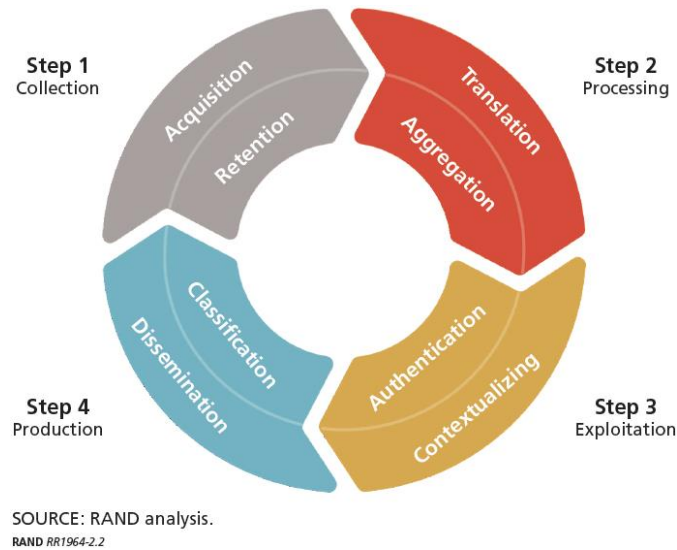


Figure 1: OSINT Operations Cycle (Williams & Blum, 2018, p14)

Rosenback & Peritz (2009, p11) present the intelligence cycle: Planning and Direction > Collection > Processing > Analysis and Production > Dissemination.

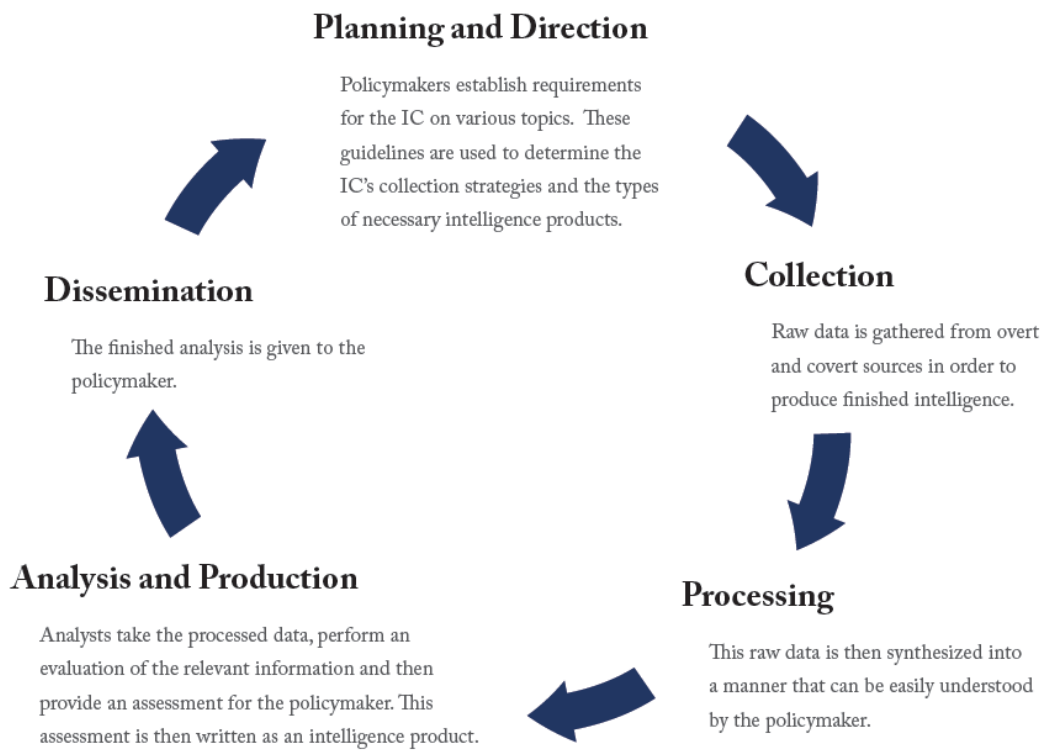


Figure 2: The Intelligence Cycle (Rosenback & Peritz, 2009, p11)

Law enforcement uses OSINT to assist in criminal investigations, such as online child solicitation, child abduction, kidnapping, cold-case homicide, terrorist threats, and high-level computer intrusions. (Bazzell, 2018) They can also gather data in attempts to intervene before crimes are committed, which falls under the banner of intelligence-led policing. (Koop, 2013, p655)

But the availability of OSINT does not necessarily mean that it's being used as much as it could to fight crime. Law enforcement must deal with internal bureaucracies, approval processes, and legal restraints. In addition, Goodman (2016, p466) notes that at the state, local, and federal levels, law enforcement officials find themselves chronically overwhelmed and understaffed as evidenced by the explosive growth in online crime.

Another caveat is the issue of falsification, as described by Bayerl & Akhgar (2015). As users in general become more concerned about online privacy and surveillance, it is not uncommon to falsify personal information, such as name and email addresses. The authors argue that, as a reaction to online surveillance, these falsification tendencies "could threaten the integrity of the very data they [law enforcement] rely on." (p68)

OSINT is widely used in the private sector, particularly in the field of advertising. In fact, the modern advertising industry is based on collecting user data. They typically use cookies, or persistent identifiers, in web browsers to track users' activities. Cookies were originally intended to help users surf the web more easily, however, the use of the "third-party" cookie enables companies to track users across many different sites. "This has evolved into a shockingly extensive, robust, and profitable surveillance architecture." (Schneier, 2015, p56) The data broker industry collects personal data from cookies and other sources, and then sells it to companies that want to know more about you.

A leader in the data broker industry is Acxiom Corporation of Little Rock, Arkansas. According to Goodman (2016, p83), Acxiom operates twenty-three thousand computer servers that collect, collate, and analyze more than 50 trillion unique data transactions every year. They have amassed profiles on over 700 million consumers worldwide. Their goal is to understand, with extreme precision, your behavioral patterns, and then package that data for sale to other companies. Companies can then use that information to craft highly targeted marketing initiatives.

The criminal element also uses OSINT to support their goals. When attacking a computer network, hackers follow a general pattern. Hutchins, Cloppert, & Amin (2011) describe the Intrusion Kill Chain, which is a “systematic process to target and engage an adversary to create desired effects.” (p4) The phases of the kill chain are:

- Reconnaissance
- Weaponization
- Delivery
- Exploitation
- Installation
- Command and Control
- Actions on Objectives (p4-5)

During the first step, Reconnaissance, the malicious actor researches the target in as much detail as necessary, using a variety of openly available sources of information such as conference proceedings, email lists, social networks, and even job postings. One popular tool for analyzing

networks is Shodan. The hacker uses the information gathered during the Reconnaissance phase to design an attack specifically for that target.

Khanna et. al (2016, p460) (2016) discuss the threat of “doxing,” which is the act of collecting publicly available personal information on an individual or organization typically for the purpose of harassing, blackmailing, or hacking purposes. The motivation is to gain revenge, financial benefit, or other cause. Doxing is considered an advanced persistent threat. (p460)

OSINT is particularly useful in spear fishing attacks, which target a specific individual within an organization. In those attacks, the attacker gathers intelligence on a person that can be exploited as part of a larger goal.

Another twist on the spear fishing attack is when an attacker creates a profile of an ideal victim, and then uses OSINT to identify and exploit potential victims. For example, in one scenario (which occurred to the author’s father), an attacker identified a retiree with a sizeable nest egg, poor computer skills, and a grandchild currently attending a private college. The attacker then initiated a phone scam in which he pretended to be that grandson in a dire financial emergency. The attacker provided sufficient detail to fool the victim into sending several thousand dollars to an out-of-state account.

OSINT in Ethical Hacking

The ethical hacker attempts to break into a system just as a criminal would, but with the approval of the client and with pre-defined limits and boundaries. “To find and fix the vulnerabilities in a computer system or network, you sometimes have to think like a criminal and use the same tactics, tools, and processes they may employ.” (Walker, 2014, p26) The EC-

Council identified five phases of hacking, which vary slightly from the Intrusion Kill Chain introduced earlier. The phases are:

- Reconnaissance
- Scanning and Enumeration
- Gaining Access
- Maintaining Access
- Covering Tracks

During the Reconnaissance phase, the hacker gathers information on the target using OSINT resources and techniques. As such, this phase is generally completely undetectable to the target. According to Walker, “Most of your vulnerability research will come down to a lot of reading, and most of that reading will come from websites devoted to informing the security crowd what’s out there.” (2014, p38)

OSINT Tools

Williams & Blum (2018) state that, “Although the tools for OSINT collection are evolving on a nearly daily basis, the methods used by the tools themselves change less dramatically.” (p23) They present a rubric from evaluating tools, based on the type of analytic methods they employ. These methods are:

- **Lexical Analysis:** aggregating and analyzing large volumes of text collected on the Internet. A simple example of lexical analysis is to identify frequently searched terms on Google. More sophisticated tools attempt to infer information about people engaging in social media, such as their demographic characteristics. Cutting-edge tools use natural

language processing and machine learning to attempt to identify shifts in ideology or viewpoints over time.

- **Social Network Analysis:** examining the connections between individuals to understand the larger network of connected actors. According to Williams & Blum (2018), Twitter is arguably the most prominent social media tool used by social network analysts. It “allows observers to see who interacts with whom, who is connected to whom—and through whom—and the quality and quantity of those interactions.” (p30)
- **Geospatial Analysis:** identifying and tracking the physical location of a target or set of targets at a given point in time. Examples include leveraging the “geotagging” features available in social media platforms such as Twitter, Facebook, Instagram, and Tumblr, as well as using the powerful open-source tools Google Earth and Google Maps. (p23-35)

OSINT tools are abundant. Bazzell (2018) provides an extensive list of OSINT tools that he finds useful for security research. He divides the tools into the following categories:

Search Engines	Social Networks	Online Communities
Email Addresses	User Names	People Search Engines
Telephone Numbers	Online Maps	Documents
Photographs	Videos	Domain Names
IP Addresses	Government Records	

Table 1: Categories of OSINT Tools (Bazell, 2018)

Instead of listing individual online resources, which are numerous, here are two sites that provide curated lists of OSINT resources:

- <https://osintframework.com>

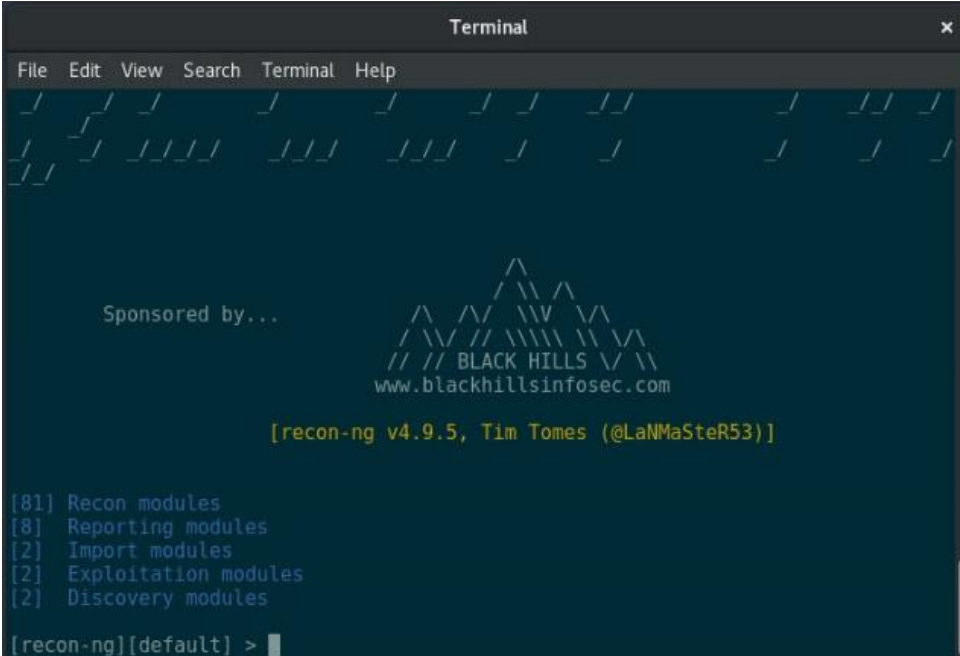
- <https://inteltechniques.com>

The following sections introduce three popular tools, Recon-ng, Maltego, and Shodan, that are used to gather and exploit OSINT.

Recon-ng

Recon-ng is a full-featured web reconnaissance framework written in Python. It provides a powerful framework in which OSINT research can be conducted quickly and thoroughly.

(Bazzell, 2018, p419-430)



```
Terminal
File Edit View Search Terminal Help
Sponsored by...
BLACK HILLS
www.blackhillsinfosec.com
[recon-ng v4.9.5, Tim Tomes (@LaNMaSteR53)]
[81] Recon modules
[8] Reporting modules
[2] Import modules
[2] Exploitation modules
[2] Discovery modules
[recon-ng][default] >
```

Figure 3: Recon-ng

Maltego

Maltego is an open-source tool that gathers information from open sources, organizes the data collected, and provides a visual representation to help determine the relationships between people, groups, and affiliations. (Khanna et. al, 2016, p460)

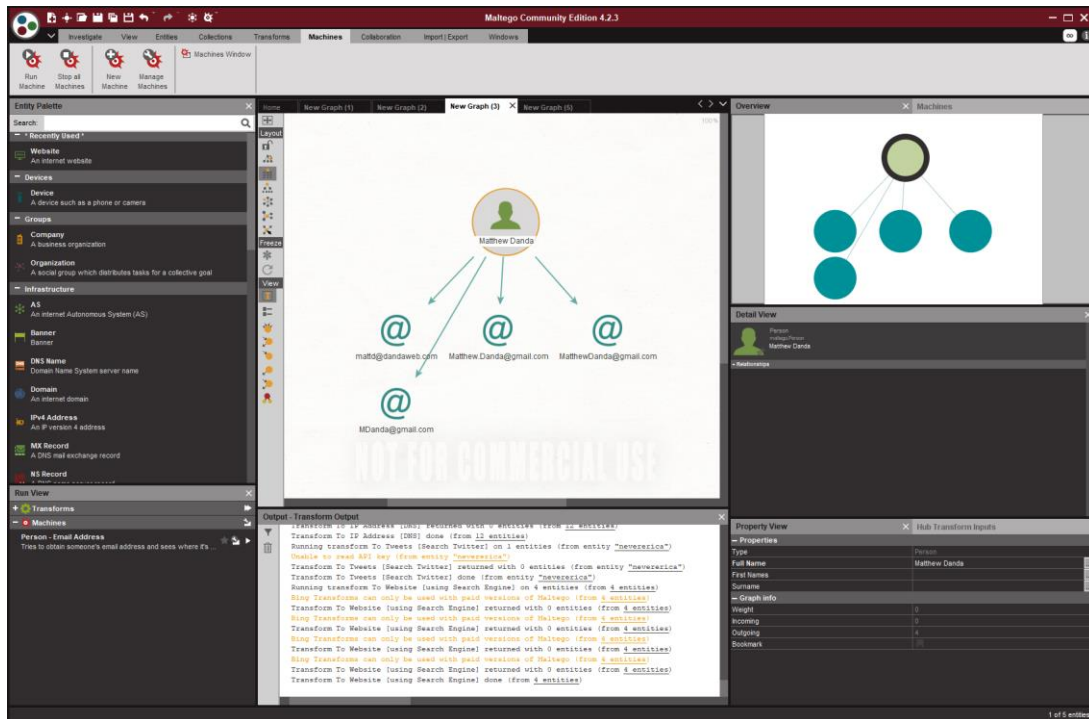


Figure 4: Maltego

Shodan

Shodan is a specialized search engine designed to find computers and devices on networks, including servers, routers, online storage devices, surveillance cameras, webcams, and VOIP systems. Shodan works by indexing banners, which are metadata that a device sends back to a client. While network administrators can use Shodan to identify vulnerabilities on their network, criminals can use Shodan to illegally access networks and other devices. (Bazzell, 2018, p343-344)

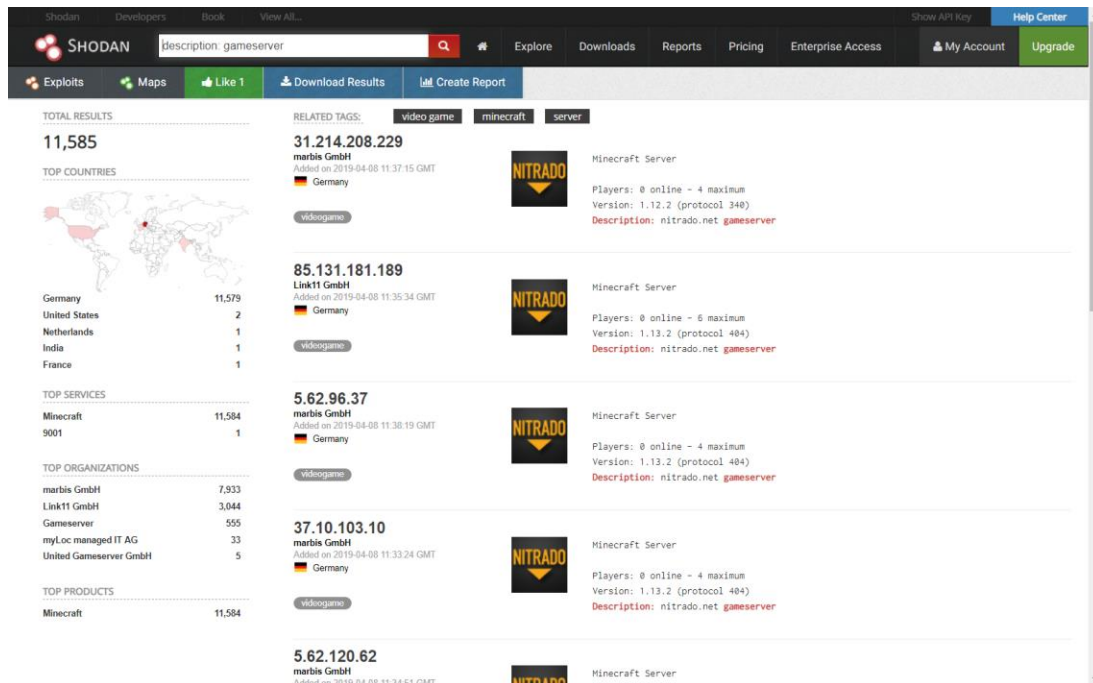


Figure 5: Shodan

Conclusion

OSINT can trace its roots to World War Two as part of the intelligence-gathering process utilized by various nation-states. However, in modern times, OSINT has emerged as a powerful resource for business, government, and criminal activity. Much of the power of OSINT stems from the unintended consequences of the internet, cloud-based data storage, and ubiquitous data-collection devices such as smart phones. Modern society is still grappling with its effects as new and creative uses for OSINT emerge. As such, OSINT must be closely monitored from academic, business, regulatory, criminal, and—perhaps above all—human rights perspectives. It is a pervasive, rapidly evolving entity with unknown levels of constructive and destructive powers. We do not know the ultimate fate of OSINT; its impact on society is still being defined.

References

- Bayerl, P., and Akhgar, B. (2015). Surveillance and Falsification Implications for Open Source Intelligence Investigations. *Communications of the ACM*. 58, 8, 62-69.
- Bazzell, M. (2018). *Open Source Intelligence Techniques, 6th Edition*. CreateSpace Independent Publishing Platform.
- Best, R., and Cumming, A. (2007). Open Source Intelligence (OSINT): Issues for Congress. *Congressional Research Service*. Order Code RL34270.
- Eijkman, Q., and Weggemans, D. (2012). Open source intelligence and privacy dilemmas: Is it time to reassess state accountability? *Security and Human Rights* 2012 no. 4.
- Glassman, M., Ju Kang, M. (2011). Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT). *Computers in Human Behavior* 28 (2012) 673-682.
- Goodman, M. (2016). *Future Crimes*. New York, NY: Anchor Books.
- Hutchins, E. M. , Cloppert, M. J., and Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Lead. Issues Inf. Warf. Secur. Res.*, vol. 1, p. 80.
- Kanakaris, V., Tzovelekis, K., and Bandekas, D. V. (2018). Impact of AnonStalk (Anonymous Stalking) on users of Social Media: a Case Study. *Journal of Engineering Science and Technology Review*, 11 (2) (2018) 126-134.
- Khanna, P., Zavarsky, P., and Lindskog, D. (2016). Experimental Analysis of Tools Used for Doxing and Proposed New Transforms to Help Organizations Protect against Doxing

- Attacks. *The 2nd International Workshop on Future Information Security, Privacy and Forensics for Complex systems* (FISP-2016).
- Koops, B.J. (2013). Police investigations in Internet open sources: Procedural-law issues. *Computer Law & Security Review* 29 (2013) 654-665.
- Krebs, B. (April 2018). When Identity Thieves Hack Your Accountant. Retrieved from: <https://krebsonsecurity.com/2018/04/when-identity-thieves-hack-your-accountant/>
- Kris, D. (March 21, 2017). The CIA's New Guidelines Governing Publicly Available Information. LAWFARE. Retrieved from: <https://www.lawfareblog.com/cias-new-guidelines-governing-publicly-available-information>
- Pellet, H., Shiaeles, S., and Stavrou, S. (2019). Localising social network users and profiling their movement. *Computers & Security* 81 (2019) 49-57.
- Rosenback, E., and Peritz, A. (2009). *Confrontation or Collaboration? Congress and the Intelligence Community*. Cambridge, Mass: The Belfer Center, Harvard University.
- Schneier, B. (2015). *Data and Goliath*. New York, NY: W.W. Norton & Company, Inc.
- Solove, D., and Schwartz, P. (2015). *Privacy, Law Enforcement, and National Security*. New York, NY: Wolters Kluwer.
- Walker, M. (2014). *CEH Certified Ethical Hacker All-in-One Exam Guide, Second Edition*. New York, NY: McGraw-Hill Education.
- Williams, H., and Blum, I. (2018). *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*. Santa Monica, CA: RAND Corporation.