THE REAWAKENING OF OPEN-SOURCE INTELLIGENCE

By Ms. Corrine Geiger

Sometimes the critical key to unlock the whole conundrum is right there under your nose….You have to know what to look for and how to recognize it when you see it.

—LTG (Retired) Samuel V. Wilson
Former Director, Defense Intelligence Agency

Introduction

Throughout history, people have made predictions that have proven to be wrong. For example, in the 1990s, they said that the internet was a passing fad, and 100 years earlier, that everything that could be invented had already been invented.[1] And more recently, many said that open-source intelligence (OSINT) was just gray literature.[2] However, these three areas are now an ever-present part of modern life, driven by mobile communications, portable electronic devices, social media, virtual workspaces, and various other technologies in the field of data. These technologies have modernized technical capabilities able to respond to oceans of data, transforming a new OSINT for modern warfare.

As of March 2020, approximately 2.5 exabytes of data were generated on the internet daily.[3] This adds up to volumes of valuable data and publicly available information (PAI) in the cyberspace domain. It comes in near real time, and an experienced professional with the right skillset can navigate through the overcrowded terrain within the cyberspace domain to track down tailored information to a specific problem set, fill intelligence gaps, and tip and cue other intelligence disciplines and collection efforts.

Perhaps most importantly, PAI gathered from the cyberspace domain can provide indications and warnings to more sensitive intelligence collection methods and warfighters. However, the swift resurgence of OSINT and the mass influx of information caused a great deal of confusion among consumers of the discipline. As a result, OSINT is often misunderstood and underutilized. This article defines and differentiates PAI and OSINT. It also discusses the use of PAI for all intelligence disciplines disparate to OSINT and OSINT's reemergence as a paramount intelligence discipline.

What Is the Difference between OSINT and PAI?

A common question adding to the perplexity of OSINT is, "What's the difference between OSINT and PAI?" There remains much bewilderment on this topic across the Department of Defense and intelligence community. OSINT and PAI are not synonymous. While OSINT cannot exist without PAI, PAI very much exists without OSINT.

PAI is broadly defined as "information that has been published or broadcast for public consumption, is available on request to the public, is accessible on-line or otherwise to the public, is available to the public by subscription or purchase, could be seen or heard by any casual observer, is made available at a meeting open to the public, or is obtained by visiting any place or attending any event that is open to the public. Publicly available information includes information generally available to persons in a military community even though the military community is not open to the civilian general public."[5] PAI has wide use outside of intelligence activities, and much of the operations community relies heavily upon it. Additionally, all intelligence disciplines use information to create either foreign intelligence or counterintelligence, the least intrusively collected of which is PAI.

Terms Analogous to OSINT

There is a lack of understanding of various terms analogous to OSINT, such as open-source information, open-source research, and open-source collection. Open-source information and PAI are interchangeable, while open-source research and open-source collection are quite divergent. Open-source research consists of any use of PAI for a non-intelligence purpose, whereas open-source collection (or collection of PAI) falls within the confines of collection; "Information is collected when it is received by a Defense Intelligence Component, whether or not it is retained by the Component for intelligence or other purposes."[4]

Only when information is collected against an intelligence requirement does the information formally become intelligence. Yet, finished intelligence products are not manufactured until the information goes through a process of collection, processing, exploitation, and dissemination. This process does not exclude OSINT. By definition, OSINT is "intelligence that is produced from publicly available information and is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement."[6] Furthermore, in accordance with the Army's policy on OSINT, the conduct of OSINT activities is authorized for "Army intelligence personnel (military, civilians, and contractors) assigned, attached, detailed to, or supporting Army intelligence organizations, units, or elements with an [authorized] OSINT mission."[7]

The Role of PAI in Intelligence

PAI presents a fundamental component of all intelligence collection and is applicable to all intelligence disciplines. "Agencies within the Intelligence Community shall use the least intrusive collection techniques feasible within the United States or directed against United States persons abroad."[8] Since PAI is generally the least intrusive means of collection, the majority of collection activities will often begin with PAI collection (see Figure 1, on the next page, for examples of PAI data sources). However, Army intelligence personnel using PAI

while operating under non-OSINT intelligence collection disciplines' authorities, rules, and regulations are not conducting OSINT activities and must refer to their respective policies. For example, a geospatial intelligence (GEOINT) professional who collects and analyzes an image from a commercial mapping site available to the public is using PAI as part of the GEOINT mission. Likewise, a counterintelligence (CI) special agent who identifies a threat from a social media platform as part of a national security investigation is conducting a CI activity. OSINT activities are unlike other intelligence disciplines' use of PAI in that PAI can be used at the tactical, operational, and strategic levels of war while producing a product that is inherently shareable with our allies and partners to drive combined joint multidomain operations. However, all OSINT collection activities must remain passive.

Other disciplines, such as human intelligence (HUMINT) and signals intelligence (SIGINT), are inherently more intrusive, can conduct active collection activities, and are intrinsically classified. OSINT does not involve messaging or personal interaction; in fact, unlike other intelligence disciplines, "OSINT collectors neither own nor control the means of collection; they must rely on others to collect, edit, and publish information, which is then subsequently acquired."[9] Unlike HUMINT, SIGINT, and GEOINT, OSINT relies exclusively on others to publish information to the public on their own accord. It is not requested, elicited, or tasked.

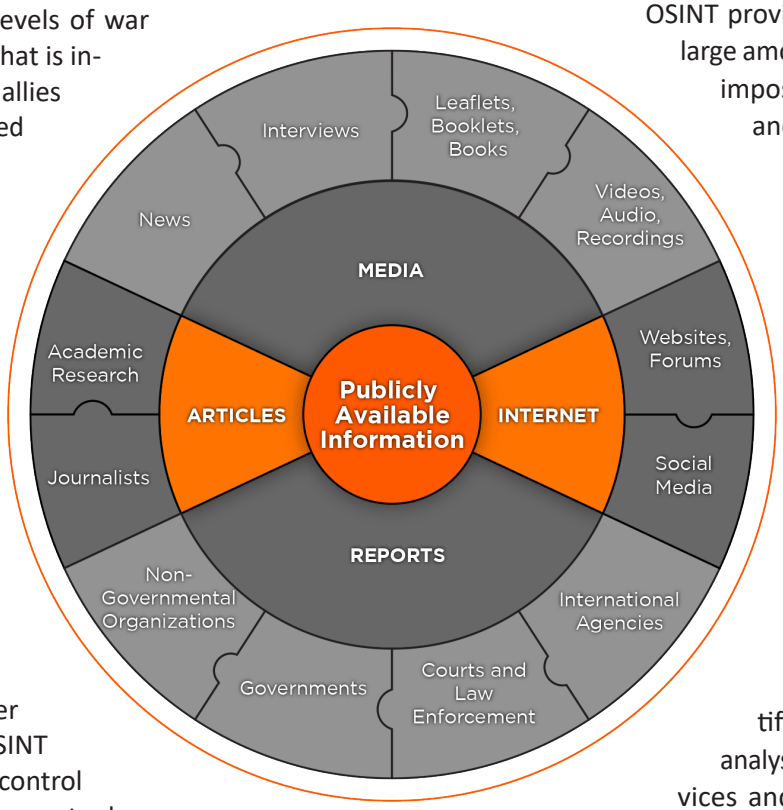

Figure 1. Publicly Available Information Data Sources

OSINT as a Discipline

A lack of control over the inordinate amount of information is precisely what makes OSINT distinct from other intelligence disciplines. OSINT collection activities extend well beyond a simple internet search. OSINT professionals must know how to triage and validate massive amounts of information. They must weed through misinformation and disinformation, propaganda, foreign malign influence, external biases, and circular reporting at levels unimaginable, without any personal interaction, while trying to determine the value of the information. OSINT professionals must also understand all the nuances of the deep and dark web, know how to navigate through them passively without standing out, and garner the most applicable information to their problem set or mission.

Approximately 10 percent of all data on the internet lies on the surface web; the rest is in the deep and dark web.[10] Yet most users are unfamiliar with these portions of the internet and lack the ability to navigate them and discover useful results. Therefore, personnel conducting OSINT activities require training on the right tactics, techniques, and procedures (TTPs). Training must be agile and constantly revised to keep up with all the social and technological changes that occur within the cyberspace domain.

OSINT Transformation

OSINT provides commanders access to large amounts of data worldwide. It is impossible to view, sort through, and verify or validate all information at that scale. As a result, the OSINT community is constantly seeking and reviewing new tools and technologies to aid with collection and prioritization. The Army focuses on building a fully integrated and unified OSINT enterprise to expand data sharing, expedite decision making at echelon, and enhance the use of cloud capabilities to include evolving artificial intelligence, semantic analysis, and machine learning services and tools. Future OSINT professionals must use technology, such as artificial intelligence and machine learning, to create and use algorithms that can shift mountains of data to answer specific intelligence requirements. Future OSINT professionals must also maintain vast familiarity of both the cyberspace domain and the information environment, have an understanding of data governance and the inner workings of the unabridged internet, and be resourced with capabilities designed to quickly sift through exabytes' worth of data.

The Army is constantly reviewing and evolving training, tools, and capabilities to meet the demands of emerging technologies so that OSINT will be "a fully operationalized intelligence discipline in support of [multidomain operations] MDO."[11] For example, as distributed ledger technologies, cryptocurrency, and metadata steganography become more prevalent, OSINT professionals will require diverse backgrounds and niche skillsets that feed OSINT's role as a single-source intelligence discipline. This is something the Army cannot accomplish on its own. The Army continues to explore partnerships with

academia and industry. It also continues to collaborate with joint and foreign partners and allies to mature the OSINT enterprise and expand OSINT's capacity.

Conclusion

Despite the fact that OSINT is derived of PAI, it is appreciably more than PAI. The magnitude of OSINT emanates well beyond that of simple information. OSINT requires a specialized skillset for safe and adequate collection, and OSINT collection activities entail tailored TTPs, processes, technologies, governance, tools, and dissemination. The internet changed the paradigm of how the world shares information. The internet is ubiquitous and will continue to drive the evolution of capabilities and ingest information at levels we cannot begin to fathom. Along with the internet, OSINT has evolved past the days of newspaper clippings, radio broadcasts, and gray literature.

The evolution of the cyberspace domain and artificial intelligence capabilities will place OSINT at the forefront of a commander's capabilities for situational awareness. It will strengthen relationships and intelligence sharing with foreign partners and allies while providing near-real-time intelligence updates and atmospherics to warfighters. OSINT is nuanced and dynamic, and is inherently tied to the cyberspace domain, which requires both a breadth and depth of information automation and data science knowledge from OSINT professionals. OSINT will never replace other intelligence disciplines, but it will enhance, broaden, and tip and cue their efforts. OSINT is often the corner and edge pieces of the puzzle that help indicate which pieces are needed to fill the center and complete the picture—it is the "source of first resort."[12] After all, "ninety percent of Intelligence comes from open sources. The other 10 percent, the clandestine work, is just the more dramatic."[13]

Epigraph

Steven Pressfield, "General Sam V. Wilson," *The Creative Process* (blog), July 2010, https://stevenpressfield.com/2010/07/general-sam-v-wilson/. This quote is from an in-depth interview Pressfield conducted with LTG Wilson.

Endnotes

1. Clifford Stoll, "The Internet? Bah! Hype Alert. Why cyberspace isn't, and will never be, nirvana," *Newsweek*, February 26, 1995, https://www.newsweek.com/clifford-stoll-said-internet-would-die-1995-566797; and Dennis Crouch, "Tracing the Quote: Everything that can be invented has been invented," Patently-O, January 6, 2011, https://patentlyo.com/patent/2011/01/tracing-the-quote-everything-that-can-be-invented-has-been-invented.html.

2. Gray literature is information produced by government agencies, academic institutions, and the for-profit sector that is not typically made available by commercial publishers. Examples of gray literature include reports, proceedings, dissertations and theses, white papers, and newsletters. "What is gray literature? How do I search for it?" Johns Hopkins University & Medicine, accessed October 19, 2021, https://welch.jhmi.edu/get-help/what-gray-literature-how-do-i-search-it.

3. "What is an Exabyte?" Wasabi, accessed October 18, 2021, https://wasabi.com/help/glossary-of-terms/exabyte-definition/. An exabyte is a multiple of a byte, which is the unit of file size for storing digital information. Since *exa* indicates multiplication by the sixth power of 1,000, an exabyte is equal to one quintillion (1,000,000,000,000,000,000) bytes or 1,000 petabytes (or 1,000,000 terabytes).

4. Department of Defense (DoD), DoD Manual 5240.01, *Procedures Governing the Conduct of DoD Intelligence Activities* (Washington, DC, August 8, 2016), 45.

5. Ibid., 53.

6. Office of the Director of National Intelligence, *U.S. National Intelligence: An Overview 2011* (Washington, DC, 2011), 54, https://www.dni.gov/files/documents/IC_Consumers_Guide_2011.pdf.

7. Secretary of the Army, Army Directive 2016-37, *U.S. Army Open-Source Intelligence Activities* (Washington, DC, 22 November 2016), 1.

8. Executive Order 12333, *United States Intelligence Activities*, as amended (December 4, 1981).

9. Mark M. Lowenthal and Robert M. Clark, *The Five Disciplines of Intelligence Collection* (Thousand Oaks, CA: CQ Press, 2016).

10. Subhendu Kumar Pani, Sanjay Kumar Singh, Lalit Garg, Ram Bilas Pachori, and Xiaobo Zhang, eds., *Intelligent Data Analytics for Terror Threat Prediction* (Hoboken, NJ: John Wiley & Sons, 2021), 309.

11. Dennis A. Eger (Defense Intelligence Senior Level, Senior Army Open-Source Intelligence Advisor), during Army Open-Source Intelligence strategy working group.

12. Former ambassador John D. Negroponte, first Director of National Intelligence, quoted in Pani et al., *Intelligent Data Analytics*, 14.

13. Pressfield, "General Sam V. Wilson."

References

DoD. DoD Instruction 3115.12. *Open Source Intelligence (OSINT)*. Washington, DC, August 24, 2010, incorporating change 2, July 16, 2020.

DoD. DoD Directive 3115.18. *DoD Access to and Use of Publicly Available Information (PAI)*. Washington, DC, June 11, 2019, incorporating change 1, August 20, 2020.

Ms. Corrine Geiger is the open-source intelligence policy advisor for the Headquarters, Department of the Army, Deputy Chief of Staff, G-2. She has more than 19 years in the Department of Defense and more than 11 years in the intelligence community. She serves in the U.S. Army Reserve as an intelligence professional, previously served as a defense contractor for both the Army and the Under Secretary of Defense for Intelligence and Security, and recently joined the ranks of Civil Service as a Department of the Army Civilian. Ms. Geiger deployed multiple times to Iraq and Afghanistan in her military capacity, and she continues to support ongoing missions as an Army Reservist. She holds a bachelor's degree in social sciences from Ashford University and is currently pursuing a master's of applied intelligence at Georgetown University.