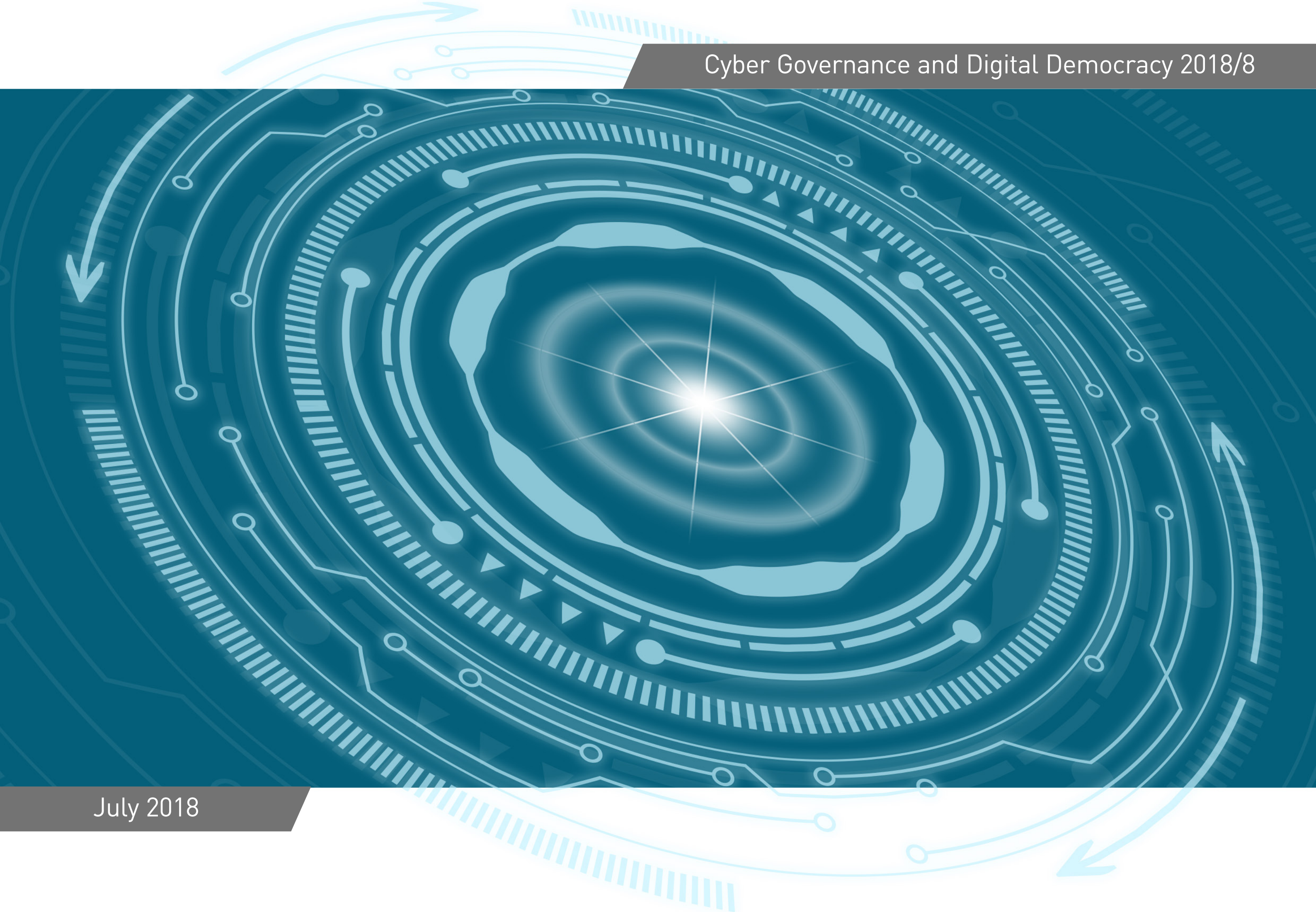
Centre for Economics
and Foreign Policy Studies

edam

Cyber Governance and Digital Democracy 2018/8



July 2018

Digital Open Source Intelligence and International Security: A Primer

H. Akın Ünver | EDAM, Oxford CTGA & Kadir Has University

Digital Open Source Intelligence and International Security: A Primer

H. Akın Ünver | EDAM, Oxford CTGA & Kadir Has Üniversitesi

Intelligence is a key and continually changing practice of statecraft. While this practice has historically been dominated by the states, merchants, and the clergy, late-20th century has witnessed the privatization of intelligence and surveillance equipment and broadening of the concept of intelligence. Today, Internet, social media, smartphones and data analytics have all contributed to the greater exposure and dissemination of critical information about emergencies and crisis events, thereby contributing to the faster travelling of news, secrets and leaks. Broadly speaking, intelligence is the practice of methodical collection and analysis of critical information for the purposes of security, or advantage. Although used synonymously with espionage, or covert operations, intelligence is mostly focused on the methodical collection, processing and analysis of information that is available and 'out there', rather than using clandestine methods to gain such information through stealing. This drive towards the collection of more and better information has been the founding block of national security, well-evidenced in successive political treatises of statecraft, since the oft-quoted 13th chapter of the Sun Tzu's 'Art of War' - The Use of Spies: '*Thus, what enables the wise sovereign and the good general to strike and conquer, and achieve things beyond the reach of ordinary men, is foreknowledge.*'[1]

The traditional understanding of intelligence is the methodical collection of high-value information in a way that yields comparative advantage to decision makers.[2] Such information can be on a foreign country's capabilities, general global events, or a country's domestic affairs. While most people tend to equate intelligence with military or security affairs, this is a very narrow definition that omits the value of intelligence in trade, finance, culture and educational affairs to render longer-term advantage during peace time. Although this traditional definition of intelligence didn't become obsolete, it was broadened through the advances in technology and more importantly, through the wide availability of such technology to wider audiences.[3] Through history, mastery of intelligence required mastery of both technology and the study of human behavior, both of which eventually rendered intelligence as a force multiplier of other functions (military, political, economic). In addition to its traditional function of enabling less miscalculated decisions, the audience of modern intelligence is growing beyond state or corporation leadership, and is expanding to the public. It is no longer a mere warning mechanism, but also a know-how reservoir and improvisation pool to resolve matters in times of unexpected crises.[4]

[1] Sun Tzu, The Art Of War (Sterling Publishers Pvt. Ltd, 2005), 92

[2] Loch K. Johnson (ed.), The Oxford Handbook of National Security Intelligence (Oxford University Press, 2010), 4.

[3] Johnson, 229.

[4] Robert Dover, Michael S. Goodman, and Claudia Hillebrand, eds., Routledge Companion to Intelligence Studies (Routledge, 2013), 51

Despite being one of the most exciting fields of inquiry in diplomacy, security and politics, the study of intelligence has consistently been difficult due to the secretive nature of the practice. Methodical information collection, establishment and maintenance of collection networks and a reliable 'information pipeline' have been some of the most crucial areas of security, without a matching scientific and scholarly rigor.[5] This was mostly due to the unavailability of historical intelligence records, or study data beyond a narrow intelligence community. However, the field has gradually opened to civilian scholarly expertise mainly in the United States, towards the end of the Cold War. This owed largely to the 1980s declassification of World War 2 intelligence files in the US and the UK, the most significant of which belonged to the Office of Strategic Services (OSS) and British signals intelligence files.[6] Previously only able to work with a small collection of cleared documents, civilian intelligence scholars now had a far larger data pool to work with. With this data availability came some of the first theories on the changing function of intelligence in national security and how it could adapt to changing technologies and communication methods.

Broadly speaking, intelligence implies four main processes. The first is collection; primarily, a state's capacity to reach, sort and collect meaningful, high-value information related to security and/or comparative policy advantage. While historically, intelligence collection capacity overwhelmingly required a wide human reach and physical access network, with 20th century, it also began to heavily include technological capacity and continuous adaptation to technical advances in communication and informatics. The second process is transmission, which involves the establishment and diversification of reliable channels of critical information flows from the target area, back to the intelligence core and from there, across domestic security institutions. Intelligence transmission requires both a highly qualified human trust network that forms an information extraction and delivery chain from the ground to the agency, as well as digital transmission structures that enable a fast delivery of digital intelligence. In intelligence types that deal with digital data - imagery, audio, text - transmission requires high levels of encryption and decryption to secure storage and transfers of such data. Third is awareness, which implies the intelligence community's understanding of the decision-makers' needs and decision-makers' understanding of the value of intelligence in key decision environments. In organizational cultures where the priorities of the intelligence community and the decision-making cohort are mismatched, or the political leadership doesn't trust the intelligence community, the awareness component is jeopardized, preventing efficient processing and transmission of key intelligence in crisis scenarios. Finally, agencies have to have the ability of 'selective deception', where it can reliably mislead competitors into wrong or missing information. This is necessary to retain comparative advantage against other intelligence competitors, by consistently distracting them into wasting resources and time on the ground.[7]

Intelligence also varies across cultures, since countries have different threat perception, information seeking and secrecy processing dynamics. To that end, intelligence should not be thought of as a monolithic and standard practice; rather, there are politically and culturally contingent ways of maximising decision-making advantage using a multitude of information gathering mechanisms. A primary determinant of intelligence culture is regime type,[8] where democracies, hybrid states and authoritarian governments process and manage information through different bureaucratic mechanisms, as well as legal and legislative oversight mechanisms.[9] In addition, democratic intelligence services tend to have greater autonomy compared to those of authoritarian states, and also tend to have a more merit-based recruitment and promotion scheme, allowing such agencies to act with greater legitimacy and a more diverse skillset against a multitude of threats. Strong oversight mechanisms also tend to minimize corruption, resource waste and mismanagement – allowing democratically-checked intelligence agencies to enjoy greater political legitimacy domestically.[10] Furthermore, authoritarian

[5] Dover, Goodman, and Hillebrand, 71.

[6] Dover, Goodman, and Hillebrand, 88.

[7] Dover, Goodman, and Hillebrand, 71–83; Johnson, The Oxford Handbook of National Security Intelligence, 113–19.

[8] Montgomery McFate, "The Military Utility of Understanding Adversary Culture" (Arlington, VA: DTIC, Office of Naval Research, January 2005), http://www.dtic.mil/docs/citations/ADA479862.

[9] Philip H. J. Davies, "Intelligence Culture and Intelligence Failure in Britain and the United States," Cambridge Review of International Affairs 17, no. 3 (October 1, 2004): 495–520, https://doi.org/10.1080/0955757042000298188.

[10] Mikael Wigell, "Mapping 'Hybrid Regimes': Regime Types and Concepts in Comparative Politics," Democratization 15, no. 2 (April 1, 2008): 230–50, https://doi.org/10.1080/13510340701846319.

states tend to inflate domestic and foreign threats, forcing wasteful intelligence agencies to spread too thin across multiple, obscure information fronts. Another determinant is institutional history and culture.[11] The intelligence practices and territorial awareness of post-imperial states (i.e. states that were once at the core of a former empire) and those that aren't, are markedly different. Inheriting a longer tradition of intelligence, such post-imperial states tend to operate across a wider territorial space, usually in the current states of their former imperial territories.[12] Finally, proximity to active conflict is crucial. States that are fighting, or adjacent to an active ongoing domestic conflict, operate on a different institutional culture compared to states that don't. Most organizational and bureaucratic models of intelligence differ according to the country's engagement with active or frozen conflicts, and/or participation in foreign peace operations.

Intelligence disciplines are roughly divided into six primary schools:

HUMINT (human intelligence): As the oldest (and up until late-19th century, the only) school in intelligence, HUMINT makes up the bulk of intelligence in history. Roughly, it relies on verbal and non-verbal communicative relations, networks and interactions between, or concerning individuals of political, military, economic or cultural importance. Psychology, cognitive mapping, sociology, anthropology and humanities are some of the key tools of the HUMINT community to understand, extract and contextualize critical security events and processes in foreign countries. Not only ambassadors, military attaches or state officials, but also traders, tourists and students have also served as a cultural and national exchange points of HUMINT throughout history. HUMINT is also by no means at the monopoly of states. Private companies, banks, research laboratories and technology firms too, engage in regular HUMINT operations (covert or overt) to achieve financial or scientific/technical advantage against their rivals.[13]

GEOINT (geospatial intelligence): Although aspects of geography (weather, terrain, waterways) have always been important variables in intelligence analysis, GEOINT has specifically benefited from the advent of real-time (or close enough) aerial imagery provided by satellites, unmanned aerial vehicles (UAVs), light detection and ranging (LIDAR) and surveillance aircraft. GEOINT provides static, or time-frequency image analysis to track and monitor human activity on a selected geographical area, as well as resources and sub-terrain conditions. Although geospatial data was previously at the intersection of MASINT and SIGINT, the availability of dedicated geospatial tools has led to the creation of the National Geospatial Agency (NGA). Today, commercially available high-resolution imagery provided by companies such as Planet Labs, Terra Bella, BlackSky Global, Orthecast or XpressSAR, have all enabled businesses, aid agencies, and a range of non-state actors to acquire GEOINT capabilities.[14]

MASINT (measurement and signature intelligence): An umbrella term for a wide array of high-technology detection tools to measure acoustic, radio frequency, radiation, chemical/biological, spectroscopic and infrared signature, MASINT is focused on collecting metric, angular, spatial and modular data through remote-sensing methods. Prior to 1991, most MASINT systems contained embedded templates and libraries of signatures to help human-assisted automated detection. Today, with the help of artificial intelligence, machine learning and big data libraries of signature detection, most MASINT systems have grown autonomous to conduct live surveillance without the

[11] Jessica L. Weeks, "Autocratic Audience Costs: Regime Type and Signaling Resolve," International Organization 62, no. 1 (January 2008): 35–64, https://doi.org/10.1017/S0020818308080028.

[12] Jeffrey W. Legro, "Culture and Preferences in the International Cooperation Two-Step," American Political Science Review 90, no. 1 (March 1996): 118–37, https://doi.org/10.2307/2082802.

[13] Jacqueline R. Evans et al., "Criminal versus HUMINT Interrogations: The Importance of Psychological Science to Improving Interrogative Practice," The Journal of Psychiatry & Law 38, no. 1–2 (March 1, 2010): 215–49, https://doi.org/10.1177/009318531003800110; Montgomery McFate and Steve Fondacaro, "Cultural Knowledge and Common Sense," Anthropology Today 24, no. 1 (February 1, 2008): 27–27, https://doi.org/10.1111/j.1467-8322.2008.00562.x.

[14] Todd S. Bacastow and Dennis Bellafiore, "Redefining Geospatial Intelligence," American Intelligence Journal 27, no. 1 (2009): 38–40; Andy Sanchez, "Leveraging Geospatial Intelligence (GEOINT) in Mission Command" (Arlington, VA: DTIC, Office of Naval Research, March 21, 2009), http://www.dtic.mil/docs/citations/ADA506270.

assistance of a human operator. Today, MASINT can be used in a wide array of information environments, from the detection of missiles, aircraft, or drones, to disaster relief, refugee aid monitoring, and natural resource - industrial output measurement.[15]

FININT (financial intelligence): With its professional motto 'follow the money', FININT is the discipline of tracking financial transactions to infer adversaries' capabilities, intentions and networks. Focusing on terrorist financing, tax evasion and money laundering, or arms trade, FININT is primarily interested in how adversaries fund their operations and assets, as well as mapping the intermediary institutions and/or persons involved in these operations. FININT is one of the most diverse schools of discipline, serving multiple branches of a government, and also one that isn't necessarily tied to security or crisis decision-making. Longer term trends that don't require a response under time or information constraints, and can be accessed through open sources, such as economic growth, industrial production, accounting policy and econometric data, are under the jurisdiction of FININT.[16]

SIGINT (signals intelligence): Although smoke, pigeons, light or semaphore signals were used as long-rage communication tools in history, the emergence of SIGINT owes mainly to the invention of telegraphy. Going as far back to 1850s as a dedicated intelligence discipline, SIGINT is primarily concerned about intercepting and processing an adversary's messages transmitted over a distance, as well as encrypting friendly communications so that they don't get intercepted by rivals. This includes tapping into communication networks and signal transmission channels for the purpose of intercepting enemy electronic communications, along with cryptographic work to handle the encryption and decryption of messages. As communication technologies have rapidly evolved through the 20th century, SIGINT has also expanded its capabilities to include TECHINT (technical intelligence), CYBINT (cyber-intelligence), and DNINT (digital network intelligence). Today, the information that lies in the vast span of the Internet, social media platforms and Internet Communication Technologies, ICTs are also under the jurisdiction of SIGINT. It is also at the forefront of current Internet-based information wars, including bots, trolls, digital spoilers and fake news.[17]

OSINT (open-source intelligence): Although an intelligence agency's capacity is primarily measured by how well it can detect and transmit critical information, its ability to understand and contextualize what is important requires the foreknowledge of what is 'out there' and easily available. To distinguish between important and redundant information, an agency must first lay the groundwork for its 'information environment'. This in turn, has to be done through developing institutional and organizational skills to cultivate and harvest information that is 'legally available in the public domain', or intelligence that is 'hidden in plain sight'. Although historically, OSINT has been driven by news and information agencies, cultural and diplomatic exchanges and socialization, it is increasingly being driven by Internet and ICT-based based technological developments. To that end, classical OSINT and digital OSINT has to be differentiated.[18]

[15] Jeffrey T. Richelson, "MASINT: The New Kid in Town," International Journal of Intelligence and CounterIntelligence 14, no. 2 (April 1, 2001): 149–92, https://doi.org/10.1080/088506001300063136; J. Dudczyk, J. Matuszewski, and M. Wnuk, "Applying the Radiated Emission to the Specific Emitter Identification," in 15th International Conference on Microwaves, Radar and Wireless Communications (IEEE Cat. No.04EX824), vol. 2, 2004, 431–434 Vol.2, https://doi.org/10.1109/MIKON.2004.1357058.

[16] Donato Masciandaro, "Financial Supervisory Unification and Financial Intelligence Units," Journal of Money Laundering Control 8, no. 4 (October 1, 2005): 354–70, https://doi.org/10.1108/13685200510620858; John Frank Thony, "Processing Financial Information in Money Laundering Matters: The Financial Intelligence Units," European Journal of Crime, Criminal Law and Criminal Justice 4 (1996): 257.

[17] Matthew M. Aid, "All Glory Is Fleeting: Sigint and the Fight Against International Terrorism," Intelligence and National Security 18, no. 4 (December 1, 2003): 72–120, https://doi.org/10.1080/02684520310001688880; Martin Rudner, "Britain Betwixt and Between: Uk SIGINT Alliance Strategy's Transatlantic and European Connections," Intelligence and National Security 19, no. 4 (December 1, 2004): 571–609, https://doi.org/10.1080/0268452042000327528.

[18] Michael Glassman and Min Ju Kang, "Intelligence in the Internet Age: The Emergence and Evolution of Open Source Intelligence (OSINT)," Computers in Human Behavior 28, no. 2 (March 1, 2012): 673–82, https://doi.org/10.1016/j.chb.2011.11.014; Robert W. Pringle, "The Limits of OSINT: Diagnosing the Soviet Media, 1985-1989," International Journal of Intelligence and CounterIntelligence 16, no. 2 (April 1, 2003): 280–89, https://doi.org/10.1080/08850600390198706.

Digital OSINT

In the words of Allen Dulles, '*A proper analysis of the intelligence obtainable by these overt, normal and aboveboard means would supply us with over 80 percent, I should estimate, of the information required for the guidance of our national policy*'.[19] Indeed, Dulles emphasizes that '*Because of its glamour and mystery, overemphasis is generally placed on what is called secret intelligence*',[20] whereas the bulk of intelligence collection and processing is usually done through 'normal methods' such as explicit diplomatic interaction, personal relationships, radio, press and a country's Diaspora abroad. The same '80% rule' is also laid down in NATO 2002 and Hulnick 2004, although for EUROPOL (European Union Agency for Law Enforcement Cooperation), the British, Swedish and Dutch ministries of defense, as well as DIA (Defense Intelligence Agency) OSINT constitutes 'at least 90%' of all intelligence activities.[21] This means that rather than the popularized and mystified practice of espionage and spying, the overwhelming majority of intelligence activities focus exclusively on harvesting open sources and finding connections and nuances where others can't.

OSINT determines the relevance and groundwork of an agency's wider functions. To that end, a proper conduct of OSINT provides two key advantages to an agency. The first is context: namely, the spectrum of events, actors and roles that determine strategic relativity (i.e. how to define a country's interests in relation to ongoing events), as well as which assets to deploy to achieve them. Without an understanding of world events, causal mechanisms between processes and explicit interests of major actors, agencies can only deal with problems reactively, without any ability to stop or manage them before they reach the nation's borders; or worse, off beyond them.[22] Second, OSINT renders other intelligence functions efficient by giving an agency an accurate understanding of what types of information are available and which ones aren't and needs dedicated focus to extract. This way, agencies can use other functions (especially more aggressive extraction mechanisms such as espionage or stealing) more sparingly, reducing the likelihood of miscalculation and escalation of tensions with another country. OSINT also decreases the costs of other intelligence functions by eliminating much of the guesswork.[23]

OSINT grew more important in influence and impact with the advances in communication and encryption technologies. The invention of the alphabet and diplomatic writing brought about the need for seals and cipher mechanisms; printing press, for officiation and modern bureaucracy; telegraph, for code-makers and code-breakers; radio, for signals interceptors (SIGINT) and computers, for high-volume encryption and decryption. The advent of the Internet, digital interconnectedness and social media platforms have all led to the growing importance of OSINT and the emergence of overlapping jurisdictional areas between other schools of intelligence, but also brought about problems of verification regarding content and news. The explosion of information and data has made life both easier and more difficult for OSINT; easier, because of the widening of the channels of communication, and hard because of a similar proliferation of junk, or misleading information. This renders OSINT's task not just collection and processing of digital data, but also developing verification and attribution mechanisms, and understanding what constitutes as junk content and what doesn't. In order for agencies to know which digital information or data type is important, they need technical infrastructure and high-quality manpower (or ability to outsource all of these functions) to grasp the Internet and its ever-changing patterns of dissemination and storage. To that end, most digital OSINT agencies have started to develop Internet studies units.[24] Furthermore, agencies not only have to compete among themselves as they historically did, but thanks to the democratization and wide availability of Internet sources to the mass public, they also have to compete with citizen analysts and private OSINT

[19] Dover, Goodman, and Hillebrand, Routledge Companion to Intelligence Studies, 125.

[20] Dover, Goodman, and Hillebrand, 125.

[21] Johnson, The Oxford Handbook of National Security Intelligence, 221.

[22] Johnson, 45.

[23] Dover, Goodman, and Hillebrand, Routledge Companion to Intelligence Studies, 14.

[24] Edward J. Appel, Cybervetting: Internet Searches for Vetting, Investigations, and Open-Source Intelligence, Second Edition (CRC Press, 2014), 157.

companies. These two new emerging intelligence actors – citizens and private analysts – are unbound by the heavy bureaucratic weight of formal intelligence agencies, and thus, can adapt to changing technicalities faster and can undertake collection, storage and analysis functions on their own initiative, for which agencies require degrees of legal legitimacy and formal authority. From counterterrorism to cybersecurity, and from WMD monitoring to protest analysis, technology companies and civilians alike tap into the same data and information types that most state OSINT agencies do. Although non-state analysts lack in financial resources of states, they make up for this shortcoming through their autonomy, speed and improvisation ability.

On top of this widening, add in the popularized variable - 'Big Data'. There are two main novelties brought about by the oft-prophesized 'Big Data Revolution': first, data storage and transmission technologies, the availability of 3G/4G data networks, mass proliferation of Wi-Fi access and cloud technologies, we are now able to produce, store and share historically unprecedented volumes of information. This both makes a given unit of data (byte) increasingly cheaper to produce, store or transmit, and also enable highly-granular social (especially personal) data to be produced and harvested. Eventually, our social and personal data has become multi-purpose; our tax and employment data for example, can be used to profile our purchasing behaviour, healthcare options, residency choice and electoral behaviour.[25] This multi-purpose social and personal data then gets even more granular through our digital behavior, in the form of Facebook friends, likes, Twitter retweets, Instagram posts, geo-located photo uploads and Snapchat videos. This allows both state and private OSINT analysts to tap into the largest, continuously-growing and extremely detailed behavioural information pool of millions of people. Finally, when considering the proliferation of 'Internet of things' (IOT) data types, from fitness watches to home appliances, this largest ever pool of social and personal data becomes enormous, yet detailed enough to profile nations in high-definition.[26] For any analyst – state or private – working on public morale, political interests, electoral choice and social forces in an adversary's society, such proliferation of data is a historically significant turn in intelligence capacity. Yet, not all states can harvest such data efficiently. For such data to be meaningfully distilled into valuable intelligence, an analyst has to possess a diverse set of competencies including computer science and data science, which is where states usually fail to catch up.

The first of many problems for state agencies is the issue of talent attraction. With Facebook, Google, Amazon and other tech companies enabling a vastly freer working environment, few (visible) hierarchies and better pay, most of the highly-qualified data analysts turn away from state service.[27] This generates a shift in the centre of gravity of digital intelligence power, from states to private companies. Second is the issue of infrastructure development, adaptation and upgrading which is problematic for the highly bureaucratic structure of the states. New hardware is always expensive and smart solutions like technology recycling (refurbishing old equipment at lower costs) or upgrade streamlining require smaller quantities and a nimbler decision-making system. The very business model of technology renders states as the trailers behind (and dependent) on technology companies.[28] Third, the growing civilianization of OSINT has created an 'information-as-resistance' movement in which digital activism implies the exposure and dissemination of state mismanagement, corruption and repression.[29] This resistance culture assumed an increasingly better-organized digital identity following with the exposure of state surveillance abuses with the Snowden revelations, Wikileaks and Chelsea Manning exposures. Although states can theoretically tap into this civilian OSINT pool, the current culture and identity of this community is mostly anti-state.[30]

[25] Westin Alan F., "Social and Political Dimensions of Privacy," Journal of Social Issues 59, no. 2 (April 29, 2003): 431–53, https://doi.org/10.1111/1540-4560.00072.

[26] Feng Chen et al., "Data Mining for the Internet of Things: Literature Review and Challenges," International Journal of Distributed Sensor Networks 11, no. 8 (August 18, 2015): 431047, https://doi.org/10.1155/2015/431047.

[27] Valerio De Stefano, "The Rise of the Just-in-Time Workforce: On-Demand Work, Crowdwork, and Labor Protection in the Gig-Economy," Comparative Labor Law & Policy Journal 37 (2016 2015): 471.

[28] Stefan Tongur and Mats Engwall, "The Business Model Dilemma of Technology Shifts," Technovation 34, no. 9 (September 1, 2014): 525–35, https://doi.org/10.1016/j.technovation.2014.02.006.

[29] Moonsun Choi, Michael Glassman, and Dean Cristol, "What It Means to Be a Citizen in the Internet Age: Development of a Reliable and Valid Digital Citizenship Scale," Computers & Education 107 (April 1, 2017): 100–112, https://doi.org/10.1016/j.compedu.2017.01.002.

[30] Zeynep Tufekci, Twitter and Tear Gas: The Power and Fragility of Networked Protest (New Haven ; London: Yale University Press, 2017).

Finally, states can potentially get hurt by OSINT, as much as they benefit from it, as OSINT is by nature, a double edged sword. A state can suffer from audience costs and public shaming from the exposure of its mismanagement and corruption, just as it tries to tap into the OSINT realm to hurt other states, or domestic opposition groups. Although civilian data leaks (voter, healthcare, purchasing history data etc.) hurt individuals, state-level data leaks hurt governments and agencies more, due to the secretive nature of most leaks.[31] This renders states larger sitting ducks in digital power parity compared to civilians (unless targeted specifically) and alters the relative power balance between the state and the society. This shift generates a security dilemma between state actors as well, as this renewed state-society power balance enables external actors to exploit and interfere with the domestic machinations of a nation. This interference can hurt powerful and weak states alike, the best example being the Russian involvement in the US elections via fake news and other publicly available news and information sources.

[31] S. Landau, "Making Sense from Snowden: What's Significant in the NSA Surveillance Revelations," IEEE Security Privacy 11, no. 4 (July 2013): 54–63, https://doi.org/10.1109/MSP.2013.90.

Types and Examples of OSINT

Although OSINT tools are rapidly evolving, most popular methods can be clustered under four main categories: linguistic/ text-based methods, geographic information systems (GIS) - remote sensing, network science, and visual forensics.

a. Linguistic and Text-Based Methods

Glossary

Natural Language Processing (NLP): Tracing its origins back to Alan Turing's 1950 article ' 'Computing Machinery and Intelligence' (from which the 'Turing test' is born), NLP is primarily interested in the interaction between human and machine language. Originally focusing on automated machine translations between human languages, NLP today is focused on the discovery of patterns within structured and unstructured, multi-linguistic and large-volumes of text, through entities, keywords, word/phrase relations and semantic/syntactic roles. NLP has paved way to more contemporary text-based methods such as automatic text summarization, machine-based sentiment analysis, entity and topic extraction and forms the foundation of modern text-mining tools.

Latent Semantic Indexing (LSI): LSI is a machine learning-based text analytics method, which learns from a sample text to identify the 'latent' concepts in multiple documents. For example, if 'artillery', 'shell' and 'bombardement' texts appear frequently in multiple documents, the system indexes these words into the same semantic context, simultaneously separating the word 'shell' from documents that contain phrases 'beach', 'sand', or 'crab'. LSI works best in large volumes of text, such as archival documents, legislation or judicial documents.

Latent Dirichlet Allocation (LDA): LDA is a text-based machine-learning method similar to LSI, although LDA clusters words into topic models by itself, rather than into folders determined by the user. LDA checks the frequency and relation of words in a text based on how frequently they are used together, and in which context.

Entity recognition-extraction: Named Entity Recognition is a process where an algorithm takes a string of text (sentence or paragraph) as input and identifies relevant nouns (people, places, and organizations) that are mentioned in that string. News and publishing houses generate large amounts of online content on a daily basis and managing them correctly is very important to get the most use of each article. Named Entity Recognition can automatically scan entire articles and reveal which are the major people, organizations, and places discussed in them. Knowing the relevant tags for each article help in automatically categorizing the articles in defined hierarchies and enable smooth content discovery.

Text corpus: A corpus is usually the main data pool for text-based OSINT methods. It is a collection of words and keywords from which statistical analyses are made. n order to make the corpora more useful for doing linguistic research, they are often subjected to a process known as annotation. An example of annotating a corpus is part-of-speech tagging, or POS-tagging, in which information about each word's part of speech (verb, noun, adjective, etc.) is added to the corpus in the form of tags.

N-Gram: In language processing, an n-gram determines the unit of analysis for the query to be searched in the corpus. If two words are searched together (i.e. 'conventional' + 'warfare', or 'terrorist' + 'attack', this query is called a bi-gram. A tri-gram on the other hand is a 3-word query that specifically searches for the combination of 'conventional' + 'submarine' + 'warfare', or 'terrorist' + 'suicide' + 'attack'.

Language and sentiment analysis has been one of the oldest practices of OSINT. Inferring leadership psychology, policy intent and organizational cohesion through speech and writing have been a core practice of historical versions of OSINT, enabling diplomats and other intermediaries to synthesize crucial information. Indeed, through the Cold War, the harvesting of newspapers, leadership statements and even scientific journals has been commonplace in countries on both sides of the conflict.[32] Furthermore, since World War I, linguistics, anthropology and area studies have grown significantly popular from an intelligence point of view, evidenced by the establishment of dedicated departments in top universities and their receipt of significant government funding.[33]

Digitization of text and the popularization of text-as-data methods in social sciences had a direct impact on linguistic OSINT analysis. Although quantitative linguistics became a popular field as far back as 1960s, mass digitization and standardization of text files through computer-based word processors, have all contributed to the significant advances in open-source harvesting such as text categorization, text clustering, entity extraction and computational summarization. Thanks to such mass digitization, entire national historical archives, political texts and memoirs have become digitized for word-processing purposes, providing linguists and content/discourse analysts with an unprecedented data size and fast processing tools. These tools have been especially valuable for Internet-based text-mining, such as websites, blogs and social media posts. In addition to the existence of 644 million websites in existence, vast volumes of social media data pour in on a daily basis, which means that an overwhelming majority of the world's text-based interactions are now searchable, sortable and measurable - some of them in real-time.

Although text-based OSINT can be done through programming standards such as Python, R, MatLab and Ruby, there are dedicated text-based OSINT applications as well. Some of the popular ones are WordStat, RapidMiner, KHCoder and NVivo that allow users to detect and visualize connections, patterns and themes in large volumes of text. In addition, natural language processing applications based on statistical topic modelling, such as Latent Dirichlet Allocation (LDA), text segmentation, Latent Semantic Analysis and Pachinko Allocation, enable a machine-learning approach for pattern detection, and sentiment analysis. Furthermore, entity-recognition and extraction applications make it far easier to catalogue, sort and process large volume of social media text data in order to do retrospective or real-time analysis.

Several promising applications of OSINT include behavioural prediction/detection, evidenced by Asghar (et. al.) work on pattern detection on Youtube comment videos to measure their level of radicalization,[34] or Hsinchun Chen's seminal work on text mining of the Dark Web,[35] and extremism networks that lie within. Singh et. al. have took this a step further and harvested Indian diplomats tweets to analyse popularity dynamics between Indian Foreign Service and Narenda Modi, giving a clear idea on diplomatic capital and support for leadership.[36] On prediction on the other hand Mueller and Rauch have used newspaper text mining to forecast imminent protests and conflicts, coming up with a clear model in using large amounts of text-as-data for forecasting purposes.[37]

[32] Johnson, The Oxford Handbook of National Security Intelligence, 144.

[33] Osamah F. Khalil, America's Dream Palace: Middle East Expertise and the Rise of the National Security State (Cambridge, Massachusetts: Harvard University Press, 2016).

[34] Muhammad Zubair Asghar et al., "Sentiment Analysis on YouTube: A Brief Survey," ArXiv 1511.09142 (November 29, 2015), http://arxiv.org/abs/1511.09142.

[35] Hsinchun Chen, Dark Web: Exploring and Data Mining the Dark Side of the Web, Integrated Series in Information Systems (New York: Springer-Verlag, 2012), //www.springer.com/gp/book/9781461415565.

[36] V. K. Singh, D. Mahata, and R. Adhikari, "Mining the Blogosphere from a Socio-Political Perspective," in 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 2010, 365–70, https://doi.org/10.1109/CISIM.2010.5643634.

[37] Hannes Mueller and Christopher Rauh, "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text," American Political Science Review, December 2017, 1–18, https://doi.org/10.1017/S0003055417000570.

b. Geospatial Intelligence and Remote Sensing Tools

Glossary

Vector and raster data: In GIS software, geographical information is stored into two main types of data. Vector data is a representation of the world using points, lines, and polygons. Vector models are useful for storing data that has discrete boundaries, such as country borders, land parcels, and streets. Raster data on the other hand, is a representation of the world as a surface divided into a regular grid of cells. Raster models are useful for storing data that varies continuously, as in an aerial photograph, a satellite image, a surface of chemical concentrations, or an elevation surface.

Basemap: A basemap provides a user with context for a map. Vector or raster data can be added to a basemap by overlaying on top of it. Basemaps contain reference information that may provide different geospatial information based on what the cartographer is trying to communicate.

Geocoding-geofencing: Geocoding is the process of transforming a description of a location—such as a pair of coordinates, an address, or a name of a place—to a location on the earth's surface. An analyst can geocode by entering one location description at a time or by providing many of them at once in a table. The resulting locations are output as geographic features with attributes, which can be used for mapping or spatial analysis. Geofencing on the other hand, is a location-based service in which an app or other software uses GPS, RFID, Wi-Fi or cellular data to trigger a pre-programmed action when a mobile device or RFID tag enters or exits a virtual boundary set up around a geographical location, known as a geofence. Depending on how a geofence is configured it can prompt mobile push notifications, trigger text messages or alerts, send targeted advertisements on social media, allow tracking on vehicle fleets, disable certain technology or deliver location-based marketing data.

GIS: A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. The key word to this technology is Geography – this means that some portion of the data is spatial. In other words, data that is in some way referenced to locations on the earth. Coupled with this data is usually tabular data known as attribute data. Attribute data can be generally defined as additional information about each of the spatial features.

LIDAR: LIDAR, which stands for Light Detection and Ranging, is a remote sensing method that uses light in the form of a pulsed laser to measure ranges (variable distances) to the Earth. These light pulses—combined with other data recorded by the airborne system— generate precise, three-dimensional information about the shape of the Earth and its surface characteristics. A LIDAR instrument principally consists of a laser, a scanner, and a specialized GPS receiver. Airplanes and helicopters are the most commonly used platforms for acquiring LIDAR data over broad areas.

Landsat: The LANDSAT program is the oldest, functional satellite imagery program, which consists of a series of optical/infrared remote sensing satellites for land observation. The program was first started by The National Aeronautics and Space Administration (NASA) in 1972, then turned over to the National Oceanic and Atmospheric Administration (NOAA) after it became operational.

Remote sensing: Remote sensing is the science of obtaining information without physically being there. For example, the 3 most common remote sensing methods is by airplane, satellite and drone. Remote sensing instruments are of two primary types—active and passive. Active sensors, provide their own source of energy to illuminate the objects they observe. An active sensor emits radiation in the direction of the target to be investigated. The sensor then detects and measures the radiation that is reflected or backscattered from the target. Passive sensors, on the other hand, detect natural energy (radiation) that is emitted or reflected by the object or scene being observed. Reflected sunlight is the most common source of radiation measured by passive sensors.

Like language, cartography too, is an old school of intelligence and strategic analysis, working primarily on geopolitical and geographical variables, as well as the political impact of borders and terrain. The combination of geographic information systems – or GIS – and Internet-based location data (check-ins, location designations) has allowed analysts to harness a wider range of social and spatial dynamics of human behaviour, including mobilization, mass movement and conflicts.[38] With the additional variables of altitude, topography, elevation, resources, transportation and infrastructure, small and large-scale human behaviour can be analysed and mapped into meaningful patterns through the use of geospatial intelligence – or GEOINT.[39] Although there are dedicated GIS platforms for this kind of analysis – ArcGis, QGis – programming platforms such as Python and R (even Excel) also have GIS packages, or extensions to integrate mapping, geostatistics and proximity analysis. With the additional imagery power of Planet Labs, Terra Bella, BlackSky Global and XpressSAR, a multitude of layers, time-frames and granularity of geographical information can now be utilized by citizen GEOINT analysts.

In GEOINT, there are two main types of data: vector and raster. Vector data is the combination of the set of polygons and coordinates to designate a specific location or area on a map. Raster data on the other hand, include imagery, elevation models and map renders to make 3D analysis. With the increasing popularity of GIS, there are significantly more geospatial databases on the Internet. These datasets are also supplemented by LiDAR (Light Detection and Ranging), UAVs, GPS and satellites to increase the granularity and size of geographic datasets. Regardless of technique, some of the best applications of GEOINT, not only supply and visualize spatial data, but also tell a policy story or see a strategic gap where other methods can't. Harvard Humanitarian Initiative for example, is one of the earlier examples of a university-led GEOINT approach. Having been established in 1999, HHI has partnered with NGOs, UN relief agencies and refugee aid organizations to map crises and conflicts in Darfur, Sudan, Chad and Congo in close partnership with ground assets.[40] During Hurricane Katrina on the other hand, both US government and non-governmental analysts have adopted different GIS methods for relief and disaster response.[41] Ushahidi - a non-profit technology company - is another notable non-state OSINT initiative, which focuses on election monitoring, disaster relief and humanitarian aid in Haiti, Chile, Kenya and Italy. Ushahidi used a 'crowdmap' - a crowdsourced map event data platform in order to crowdsource crisis events.[42] Crowdmap was deployed in a number of protests around the world, including Occupy movements, 2011 London protests, in addition to the company's famous event monitoring of the 2007-2008 Kenyan crisis. Later on, Ushahidi provided the infrastructure for crisis event data collection based on witness accounts and was deployed to monitor the elections in Italy and India.[43]

[38] Thomas Zeitzoff, "How Social Media Is Changing Conflict," Journal of Conflict Resolution 61, no. 9 (October 1, 2017): 1970–91, https://doi.org/10.1177/0022002717721392; Seva Gunitsky, "Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability," Perspectives on Politics 13, no. 1 (March 2015): 42–54, https://doi.org/10.1017/S1537592714003120.

[39] Bacastow and Bellafiore, "Redefining Geospatial Intelligence."

[40] Steve Lohr, "In Relief Work, Online Mapping Yet to Attain Full Potential," The New York Times, March 28, 2011, sec. Business Day, https://www.nytimes.com/2011/03/28/business/28map.html.

[41] Jeffrey Gettleman, "Congo: Rapes by Civilians Rise Sharply, Study Says," The New York Times, April 14, 2010, sec. Africa, https://www.nytimes.com/2010/04/15/world/africa/15briefs-congo.html.

[42] Anand Giridharadas, "Ushahidi - Africa's Gift to Silicon Valley: How to Track a Crisis," The New York Times, March 13, 2010, sec. Week in Review, https://www.nytimes.com/2010/03/14/weekinreview/14giridharadas.html.

[43] Sarah Wheaton, "New Technology Generates Database on Spill Damage," The New York Times, May 4, 2010, sec. U.S., https://www.nytimes.com/2010/05/05/us/05brigade.html.

c. Connections and Networks

Glossary

Network nodes: In a communications network, a network node is a connection point that can receive, create, store or send data along distributed network routes. Each network node -- whether it's an endpoint for data transmissions or a redistribution point -- has either a programmed or engineered capability to recognize, process and forward transmissions to other network nodes.

Density: The density statistic represents the proportion of possible relationships in the network that are actually present. The value ranges from 0 to 1, with the lower limit corresponding to networks with no relationships and the upper limit representing networks with all possible relationships. The closer the value is to 1, the denser is the network and the more cohesive are the nodes in the network. Information in dense networks can flow more easily than information in sparse networks.

Centrality (betweenness): In network analysis, centrality designates the most important nodes in a graph, with regard to the number of connection to other nodes. In OSINT, network centrality studies usually focus on the most important, or best-connected members of a large group. In social network analysis, high-centrality figures are those that assume influencer status.

Homophily: Network homophily is a theory, which argues that similar nodes are more likely to attach to each other than dissimilar ones. In dense and large social networks, homophily measure enables an analyst to identify a community or a group within a larger population pool easily. Homophily is a key topic in network science as it can determine the speed of the diffusion of information and ideas.

Relations, groups and networks have always been popular for OSINT. Organizational leadership, political decision-making circles, terrorist inner circles have been central topics of inquiry for intelligence analysis. Classical network theory focuses on social networks among individuals (friendships, advice-seeking..) and formal contractual relationships (alliances, trade, security community). What makes network theory important to social science, politics and IR is its ability to conceptualize and theorize relations at the micro, meso and macro-levels of analysis in political processes, offering a structure to seemingly complex interactions. Accordingly, network theory stipulates that relations and internal-external pressures on those relations have the ability to affect beliefs and behaviors. Instead of adopting IR's mainstream levels of analysis approach, network theory focuses on the interactions between these levels of analyses, aiming to conceptualize how these interactions lead to policy and behavior.[44] A variety of applications such as Gephi, NetMiner and iGraph have made it easier to work with larger networks and measure them by betweenness, homophilly and centrality using quantitative methods. This enables extremism and radicalization networks easier to visualize and contextualize the role of hierarchies and influencers much clearer compared to traditional methods.[45] Computational network analysis on the other hand, expands classical network theory to far larger size and complexity levels, not only designating relations between them, but also use artificial intelligence, machine learning and neural networks approaches to automatically generate real-time changes in these relations. Today, one of the most popular uses of network analysis in digital OSINT is social media analysis; namely, follow, like and share relations between very large groups.[46] Compared to older methods, social network analysis enabled influencers and hierarchies in these systems more successfully.

[44] Johnson, The Oxford Handbook of National Security Intelligence, 26.

[45] Matt Apuzzo, "Who Will Become a Terrorist? Research Yields Few Clues," The New York Times, December 21, 2017, sec. World, https://www.nytimes.com/2016/03/28/world/europe/mystery-about-who-will-become-a-terrorist-defies-clear-answers.html.

[46] Jytte Klausen, "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq," Studies in Conflict & Terrorism 38, no. 1 (January 2, 2015): 1–22, https://doi.org/10.1080/1057610X.2014.974948.

d. Image and Video Forensics

Glossary

Artefact: An artefact is a visible distortion and visual error in a media (video, audio or image). Artefact homogeneity is a media forensics tool that measures whether a media type is manipulated by measuring the extent to which these distortions are even throughout the media. Artefact unevenness is usually associated with manipulation and doctoring, and can be noticed through machine-learning-based media forensics tools.

Digital Forensics: The area of digital media forensics is both the vocation of finding deleted or hidden data and also the mastery of the underlying technologies behind the various tools used and the ability to present scientifically valid information. Digital media forensics is a growing science that allows governments and corporations to assess the genuineness of digital evidence.

Photographic Comparison: As an image forensics tool, photographic comparison tests the genuineness or alterations across multiple versions of the same image. On the Internet, photographic comparison is required to sort through large quantities of similar images to find the original version. Especially in images related to crisis events, or photographs that have high political value, automated comparison software can be used to detect unevenness that can't be measured by the human eye.

Metadata: Media files contain properties that describe the contents of the file. These properties can be categorized as follows:
a) Media-type attributes that specify the encoding parameters, such as the encoding algorithm (media subtype), video frame size, video frame rate, audio bit rate, and audio sample rate.
b) Metadata contains descriptive information for the media content, such as title, artist, photographer, and genre. Metadata can also describe encoding parameters. It can be faster to access this information through metadata than through media-type attributes.
c) DRM properties, which contain information on usage restrictions. Currently Media Foundation does not support DRM properties through metadata, with the exception of the PKEY_DRM_IsProtected property.

Photogrammetric Analysis: Originally a tool for MASINT, photogrammetry is the science of extracting measurements from photographs. Such measurements can be exact coordinates, or distance between images on the media. Currently, OSINT analysts can conduct digital photogrammetric analysis through 2-D and 3-D images collected through satellite, drone or LIDAR imagery. Algorithms for photogrammetry typically attempt to minimize the sum of the squares of errors over the coordinates and relative displacements of the reference points.

As wifi and phone data network services became faster and cheaper, online human communication has rapidly evolved from text-based to media-based. We usually find it easier to send a voice message on Whatsapp instead of texting, or to send a photo or a video to express longer sentences and paragraphs. The same logic works for crises and emergencies. Under stress, people tend to share images and videos to document, or call for help, instead of texting and typing long messages online. To that end, although we tweet, share and blog, the increasing majority of our digital communication (especially during crises) has become media-based. While studying photographs for strategic gain or emergency communication goes back to the late-19th century, 'video intelligence' as a common practice, is mostly a post-World War 2h endavor. Today, such visual media can be digitally analysed, interpreted and used to extract key information from the ground - especially in conflict, protest or disaster areas where physical access is limited. Images and videos can be used for verification, statement, propaganda and counter-propaganda purposes on the battlefield, or in crisis episodes; they can be shared as an evidence of relations, interests and capabilities. Due to the value of emergency media for OSINT, this is also one of the most vulnerable areas for manipulation and forgery. Images and videos alike can be faked, doctored, and old media can be shared as new. This in turn allows state and non-state actors to mislead, distract and intimidate their rivals during emergencies.

Several private initiatives have embarked on a dedicated study of web-based images and videos to form a crowdsourced OSINT network, the most famous being Bellingcat – the online investigation platform. Bellingcat has published several tutorials on how to conduct media-based OSINT, and some of its famous investigations include Russian troop movements, Syrian chemical weapons use verification and protestor-riot police dynamics in a number of incidents.[47] Another example is Forensic Architecture - an academic-activist platform headquartered at the University of London - which uses photos, videos and aerial imagery to reconstruct poorly documented incidents that contain political importance.[48] Both Bellingcat and Forensic Architecture aim to verify critical events through a methodical study of media, as well as stitching scattered visual evidence together through diverse sources in order to create evidence. Initially viewed as an enthusiast's hobby, media forensics OSINT initiatives have grown more relevant and efficient compared to state intelligence agencies, evidenced by the fact that both Bellingcat and Forensic Architecture initiatives have provided court evidence, as well as data for UN and state-led human rights reports.[49]

Crowdsourced OSINT

With so many plentiful and publicly accessible critical data types, it is quite tempting to make the case that the 'secrets are over', or that we are entering a 'post-secret' world order. Indeed, when Sean P. Larkin heralded the 'Age of Transparency' in his famous Foreign Affairs piece, he was adamant that the proliferation of commercially-available satellite imagery, drone sensing, automated crisis reports, citizen journalists and open source bloggers would render secrets meaningless.[50] His point was that due to the decreasing costs of publicly available surveillance, the costs of acquiring and protecting secrets were increasing. States' ability to create and sustain frames and narratives (ontological security) during crises, diplomatic escalations and protests has been substantially hampered by technology. Especially since the global discovery of the power of social media during key events, states had to compete with new sources of narratives and framing beyond the conventional news sources.

Global interconnectedness and the emergence of 'citizen-led reporting' has brought about a new analyst caste: crowdsourced intelligence network, aiming to harness the labor of like-minded digital activists to challenge and counter state narratives. In fact, it was the United States (DARPA) that had first tried to use crowdsourcing for intelligence analysis, through its 2009 digital exercise titled 'Network Challenged'.[51] During this crowdsourcing exercise, a multitude of challenges faced by state-led efforts in OSINT (such as fast verification, event data generation, measurement) could be better managed by a semi-autonomous network of users, working through social networking tools. This exercise has demonstrated that 'amateurs' (meaning civilians that had little or no formal background in intelligence and policy planning) were both useful and not so useful from different perspectives. Crowdsourced OSINT was definitely fast, unbound by the constraints of bureaucracy and strict policy. On the other hand, most of these OSINT enthusiasts lacked sufficient intelligence training, policy organization and coherence in preparing policy options for decision-makers. In other words, crowdsourced OSINT was deemed to be good at challenging state narratives during a focused incident (like a crisis), but lacked capacity to monitor and harvest regular, daily Internet data to designate political patterns and come up with policy suggestions.[52]

[47] Pablo Gutierrez and Paul Torpey, "How Digital Detectives Say They Proved Ukraine Attacks Came from Russia," The Guardian, February 17, 2015, sec. World news, http://www.theguardian.com/world/2015/feb/17/ukraine-russia-crossborder-attacks-satellite-evidence.

[48] Rowan Moore, "Forensic Architecture: The Detail behind the Devilry," The Observer, February 25, 2018, sec. Art and design, http://www.theguardian.com/artanddesign/2018/feb/25/forensic-architects-eyal-weizman.

[49] Dylan Collins, "A US Airstrike Which Killed 38 People Allegedly Hit a Peaceful Mosque in a Syrian Village," Business Insider, April 18, 2017, http://www.businessinsider.com/us-airstrike-allegedly-hit-a-peaceful-mosque-in-a-syrian-village-2017-4.

[50] Sean P. Larkin, "The Age of Transparency," Foreign Affairs, April 18, 2016, https://www.foreignaffairs.com/articles/world/2016-04-18/age-transparency.

[51] Mark Harris, "How A Lone Hacker Shredded the Myth of Crowdsourcing," WIRED, September 2, 2015, https://www.wired.com/2015/02/how-a-lone-hacker-shredded-the-myth-of-crowdsourcing/.

[52] Larry Greenemeier, "DARPA Verigames Crowdsourced Formal Verification (CSFV) Project," Scientific American, June 9, 2015, https://www.scientificamerican.com/citizen-science/darpa-verigames-crowdsourced-formal-verification-csfv-project/.

In addition, it is also politically hard for states to harness the power of crowdsourcing in OSINT during crisis events. Given how most digital OSINT tools became globally commonplace with 2011 following the Occupy and Arab Spring movements, the overall tone of the practice became anti-hegemonic and oppositional.[53] Most earlier forms of crowdsourced OSINT focused on steering protest crowds, organizing protest logistics and circumventing the police, or state intelligence agencies. To that end, a wide chasm emerged between state intelligence agencies that mistrust OSINT, and citizen-led analytics that mistrust the motives of the state. This mutual mistrust has so far prevented a workable model for state-led efforts in cultivating a crowdsourced OSINT environment. Since such a model unforthcoming, states and citizen-led efforts use their own tools and networks during emergencies.

Crowdsourcing involves ground-based event data producers, near-ground-based data curators and off-site, remote location data analyzers. One the earlier good examples is the Ushahidi (means 'witness' in Swahili) platform that mapped election-related violent activities in Kenya between 2007-2008. Ushahidi's event data detection performance did a better job than state intelligence actors in monitoring the conflict there, as it still is the primary data source on Kenyan election violence as of today.[54] Ushahidi later switched to a GeoCommons mapping platform to mobilize and crowdsource event data on the 2010 Haiti earthquake, considerably helping aid and relief agencies in their efforts to respond to as many incidents as possible. Ushahidi would later grow to a level of importance that the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) partnered with the platform to create a Libya Crisis Map, to gather ground data in war-stricken areas that need aid drop.[55] In one of the earlier examples of how such civilian-led relief and aid OSINT platforms could be exploited by state actors, some NATO air assets used this aid map to refine ground targets and schedule aerial bombardments.[56]

Bellingcat and LiveUAMap are two of the newer additions to crowdsourced intelligence. Bellingcat appeared from its humble beginnings in 2012 as a blog and LiveUAMap during the earlier phases of the Russian military involvement in Ukraine in 2014. Bellingcat rose to fame in 2014, when its crowdsourced analysts used open-source tools to discover which Russian unit shot down the MH17 flight in Ukraine.[57] This investigation was a turning point in OSINT, as its use of publicly available information generated stronger evidence against Russian involvement in the MH17 shooting compared to all other state-produced evidence reports. Ultimately, it was Bellingcat's report that was incorporated into the indictment at the Dutch court that was handling the investigation.[58] Later in 2014, Bellingcat would publish successive online reports on the use of cluster munitions and other internationally banned area-of-effect weapons, demonstrating how the Syrian Army was producing, transporting and deploying these banned weapons.[59] Then in 2015, Bellingcat became the first OSINT outlet that discovered the shifting drone tactics of ISIS and their invention of the grenade-dropping UAVs.[60] The initiative has since grown considerably in fame and volunteers, building up a network of crowdsourced event data producers, video and image analysts and GIS mappers around the world. The group has also begun teaching their OSINT methods to enable more citizen-led intelligence production.

LiveUAMap works slightly differently than Bellingcat. LiveUAMap harvests social media data in near-real-time in order to display and map conflict events on their interactive world map. Although the group initially started as an outlet

[53] Tufekci, Twitter and Tear Gas.

[54] Giridharadas, "Ushahidi - Africa's Gift to Silicon Valley."

[55] John D. Sutter, "Ushahidi: How to 'crowdmap' a Disaster," CNN Labs, October 25, 2010, http://www.cnn.com/2010/TECH/innovation/10/25/crowdmap.disaster.internet/index.html.

[56] Ian Traynor, "Libya: Nato Bombing of Gaddafi Forces 'Relying on Information from Rebels,'" The Guardian, May 18, 2011, sec. World news, http://www.theguardian.com/world/2011/may/18/libya-nato-bombing-benghazi-rebel-leaders.

[57] Max Fisher, "Did Ukraine Rebels Take Credit for Downing MH17?," Vox.com, July 17, 2014, https://www.vox.com/2014/7/17/5913089/did-this-ukrainian-rebel-commander-take-credit-for-shooting-down-the.

[58] Mark Gibney, "The Downing of MH17: Russian Responsibility?," Human Rights Law Review 15, no. 1 (March 1, 2015): 17, https://doi.org/10.1093/hrlr/ngu036.

[59] Martin Chulov, "Syria Attack: Nerve Agent Experts Race to Smuggle Bodies out of Douma," The Guardian, April 12, 2018, sec. World news, http://www.theguardian.com/world/2018/apr/12/syria-attack-experts-check-signs-nerve-agent.

[60] Ben Sullivan, "The Islamic State Conducted Hundreds of Drone Strikes in Less Than a Month," Motherboard, February 21, 2017, https://motherboard.vice.com/en_us/article/vvxbp9/the-islamic-state-conducted-hundreds-of-drone-strikes-in-less-than-a-month.

to monitor specifically Russian activities in Ukraine in 2014, it has grown in scope to include Syria, Iraq, and then, rest of the world. LiveUAMap is a truly crowdsourced conflict monitoring platform that uses social media data in multiple languages to create real-time alerts, as well as a database of events that go as far back to 2013. Similar initiatives are FlightTracker, which maps and displays the code information and destination of commercial, as well as government flights, TankerTracker, which tracks oil and natural gas tankers across the world's main ports, and DroneDeploy, which provides a real-time visualization of major combat and reconnaissance drones deployed by militaries and non-state actors across the world.

International Political Implications of OSINT: Democracy and Security Dilemma

In November 2017, the fitness-tracking app and gadget maker Strava, has released its users' dataset containing 13 trillion GPS location data points.[61] Initially thought of as a way to help people socialize through their fitness performance (i.e. how much, or fast they ran) by sharing personal scores on social media, the release turned out to be an operations security disaster. While these individual location data points revealed popular running routes in major cities, they also revealed unidentified military bases via soldiers' Strava tracker use. While the commercialization of drone and satellite imagery have already led to the discovery of most major military installations in the world, Strava data took this one step further: exposure of secret military installations (especially those in combat zones) and the time, date and trajectory of runners in those military bases. Although unintended, this was such a major security breach that Colonel John Thomas, a spokesperson for the US Central Command gave a statement to the Washington Post that the military was 'looking into the implications of the map'.[62]

Then in mid-March 2018 the data analytics company Cambridge Analytica's extra-judicial dealings with the Trump campaign were exposed, revealing how 50 million Facebook profiles were harvested without consent.[63] Facebook was directly involved as an active actor in the scandal, by willingly exposing 50 million profile raw data to Aleksandr Kogan, a senior Analytica data scientist, who had close contacts with Steve Bannon, who was a major leader within the Trump campaign. Kogan had built 'thisismydigitallife' – a quiz app on Facebook – which pro led an initial 270,000 Facebook users who took the quiz, without the knowledge of this data to be used in a political campaign.[64] Through network analysis methods (friends, interests, likes) Kogan was able to access 50 million users' data through this initial 270,000. More recently, a group of political scientists have used text-based machine learning methods to analyze classification patterns of US State Department cables since 1971.[65] These cables contained correspondence between the State Department and a US diplomatic mission in a foreign country. By studying the content of millions of cables, the researchers have identified which word combinations are likely to be in cables that are flagged as 'secret', 'confidential', 'limited official use', or 'unclassified'. The study has revealed that human error plays a considerable role in the misclassification of secrets, leading to the declassification of a large number of secret documents. Most critically, the study has discovered that there are uneven rules that govern whether a document is 'secret'. This, researchers argue, can both allow other states to use machine learning tools to extract secret information through declassified US archives, and also for civilian analysts to tap into the same reservoir to leak secrets to

[61] Alex Hern, "Fitness Tracking App Strava Gives Away Location of Secret US Army Bases," The Guardian, January 28, 2018, sec. Technology, http://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

[62] Andrew Liptak, "Strava's Fitness Tracker Heat Map Reveals the Location of Military Bases," The Verge, January 28, 2018, https://www.theverge.com/2018/1/28/16942626/strava-fitness-tracker-heat-map-military-base-internet-of-things-geolocation.

[63] Alvin Chang, "The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram," Vox, March 23, 2018, https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram.

[64] Carole Cadwalladr, "'I Made Steve Bannon's Psychological Warfare Tool': Meet the Data War Whistleblower," The Guardian, March 18, 2018, sec. News, http://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump.

[65] Renato Rocha Souza et al., "Using Artificial Intelligence to Identify State Secrets," ArXiv 1611.00356 (November 1, 2016), http://arxiv.org/abs/1611.00356.

the public. All three incidents demonstrate how states and civilians alike can be the victims of OSINT and neither 'side' has the real upper hand in the vast analytic ocean.

Following the Weberian logic that states are the sole legitimate wielders of organized violence, the same can be applied to the field of secrecy. States are usually thought of as the sole legitimate wielders of organized, institutional secrecy. Most voters in democracies and authoritarian systems alike, think that states should be capable of collecting and reliably processing large troves of intelligence concerning national security, as well as be able to protect those secrets from rival access. Yet, what separates democracies from authoritarian systems is the issue of intelligence oversight and safeguards against abuses of such secrecy.[66] Citizens and domestic targets are often the most vulnerable and easiest targets against such abuse, since the very counterintelligence and secrecy-accumulation tools that states use to achieve security can be used to track and suppress domestic dissent.[67] On the other hand, greater transparency and accountability saps intelligence agencies' speed and operational range of intelligence agencies, having a negative effect on national security. This produces an inherent dilemma for those seeking to achieve a middle ground between privacy and security: on the one hand, unchecked intelligence agencies do and will impair a country's democratic functioning by abusing the vast surveillance apparatus. Most policies that successfully render intelligence activities more transparent end up disabling intelligence services' effectiveness, scope and deterrence capability.[68]

The prevalent argument against oversight in democracies is that intelligence is not a 'regular' policy area that can be restrained through normal judicial and legislative means.[69] By rendering intelligence activities subject to lengthy legal and parliamentary fact-finding and supervision, countries may a) miss a critical intelligence interception, b) lose the comparative advantage of sensitive information to the intelligence services of authoritarian countries, and c) fail to prevent an attack, which will generate far more significant public backlash compared to intelligence abuse.[70] An authoritarian government that has none (or few) of these democratic constraints can become nimbler and faster over the short-term to meet the requirements of global intelligence rivalry and score an advantage against democracies – or so the primary argument goes. There are two main problems with this argument. First, as Desch,[71] and Reiter (et. al.)[72] have demonstrated, democracies too, can keep large amounts of information hidden from the public eye and can also successfully avoid oversight mechanisms. As recurring examples show, democracies are as likely as autocracies to go to unilateral or diversionary wars by misleading public opinion.[73] Secondly, there is no evidence to support the claim that oversight mechanisms or safeguards against intelligence abuse render democracies strategically less advantageous than autocracies. The general trend established in the literature (with few outliers) still remains robust: due to open information and a wider 'marketplace of ideas', democracies tend to miscalculate and misperceive less, don't fight with each other, tend to suffer less from civil wars and domestic disturbances, and end up winning most of the wars they get into.[74] So, what's the problem?

The causal mechanism between intelligence oversight and strategic disadvantage is weak at best, due to fact that it is

[66] Michael P. Colaresi, Democracy Declassified: The Secrecy Dilemma in National Security (Oxford, UK: Oxford University Press, 2014).

[67] Zachary K. Goldman and Samuel J. Rascoff, Global Intelligence Oversight: Governing Security in the Twenty-First Century (Oxford: Oxford University Press, 2016), 14.

[68] Goldman and Rascoff, 72.

[69] Daniel Baldino, ed., Democratic Oversight of Intelligence Services (Sydney: Federation Press, 2010), 3.

[70] Johnson, The Oxford Handbook of National Security Intelligence, 80.

[71] Michael C. Desch, "Democracy and Victory: Why Regime Type Hardly Matters," International Security 27, no. 2 (October 1, 2002): 5–47, https://doi.org/10.1162/016228802760987815.

[72] Dan Reiter, Allan C. Stam, and Alexander B. Downes, "Another Skirmish in the Battle over Democracies and War," International Security 34, no. 2 (September 30, 2009): 194–204, https://doi.org/10.1162/isec.2009.34.2.194.

[73] Erich Weede, "Democracy and War Involvement," Journal of Conflict Resolution 28, no. 4 (December 1, 1984): 649–64, https://doi.org/10.1177/002200278402800400; Kenneth A. Schultz, "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War," International Organization 53, no. 2 (ed 1999): 233–66, https://doi.org/10.1162/002081899550878.

[74] Johnson, The Oxford Handbook of National Security Intelligence, 167; Baldino, Democratic Oversight of Intelligence Services, 45; David Lyon, Kirstie Ball, and Kevin D. Haggerty, Routledge Handbook of Surveillance Studies (New York: Routledge, 2012), 51.

rarely safeguards that render intelligence ineffective.[75] Fast and good intelligence are two different things, as well as the fact that fast intelligence doesn't always lead to good policy. Although democracies may lose time and range with their intelligence operations through the constraints set by safeguards, they more than make up for this short-coming in two areas. First, due to intelligence safeguards and oversight mechanisms, agencies have to pass through a review system that tests the rationale, reasoning and strategic utility of surveillance practices.[76] This additional layer of oversight has a likelihood of spotting mistakes or misjudgements early on, preventing agencies to get sucked up into a costly mistake or an international incident that will lead to diplomatic escalation with another country. Second, democracies tend to be less concerned with the ideological purity of the intelligence community and more with its technical level of skill and capacity. In most authoritarian regimes, influential positions in intelligence are filled with commissar-type appointees, or relatives that have little, or insufficient operational/technical expertise.[77] In ideologically-driven intelligence agencies, where capability is a secondary consideration in appointments, fast decision-making usually ends up in costly miscalculations, offsetting the speed and range benefits of not having oversight mechanisms or safeguards. Therefore, although democracies may make slower intelligence decisions, these are usually made by a more technocratically-oriented community, with better interaction between the decision-making, judicial organs and technocrats, ultimately leading to better-formulated and less crisis-prone policies. This eventually renders democratic intelligence practices more likely to lead to good national security policy, compared to authoritarian systems.

In the same capacity, the 'intelligence dilemma' – namely, the notion that states are 'secrecy maximizing' actors that operate in a zero-sum information environment – may be less important than argued. First of all, states collect, process and store intelligence commensurate with their technical, human and bureaucratic infrastructure.[78] States cannot be intelligence-maximizing actors, simply because once they accumulate secrets beyond their infrastructure limits, they end up becoming unable to protect them against foreign spying. To that end, states are secrecy-optimizing actors that have to prioritize the type of information they spend their infrastructure on, so that they can process them meaningfully for decision-making and to protect such secrets at a pareto-optimal cost against foreign prying.

OSINT has changed this equation substantially. High-quality intelligence is no longer in the hands of a small monopoly of states and powerful corporations. Journalists, NGOs, and citizens too, now have the tools access, harvest, process and disseminate previously classified information. The marketization of intelligence - surveillance equipment, social media analytics services and the programming revolution - led to the emergence of new power sources in international intelligence competition. Hackers are old news - these non-state actors have already grown into a regular variable in strategic competition, be it independent, or state-supported. Emerging power sources in OSINT don't have to possess the coding ingenuity of hackers. Availability of commercial satellite imagery, over-the-counter drones, social media analytics platforms and a bit of free time have all contributed to the advent of the global OSINT caste, with disproportionate influence over information politics. Today, enthusiasts with modest levels of technical knowledge, less-than-basic programming ability and a keen eye for exploring digital media data can become a part of the global crowdsourced OSINT network.

Now states not only have to think of other states, or big corporations as intelligence competitors, nor hackers, but also this global network of citizen journalists, OSINT enthusiasts and civilian data analytics initiatives. This network is becoming increasingly more influential on challenging or supporting state-led information operations, propaganda and political communication warfare, often yielding major international evidence, as illustrated by the Bellingcat's MH17 flight forensics work.[79] How states should respond to the advent of digital crowdsourced OSINT is largely a regime-type question, due to the role of secrecy in state-society relations. Normally, it is expected that authoritarian

[75] Hans Born and Ms Marina Caparini, Democratic Control of Intelligence Services: Containing Rogue Elephants (Ashgate Publishing, Ltd., 2013), 4.

[76] Baldino, Democratic Oversight of Intelligence Services, 89.

[77] Johnson, The Oxford Handbook of National Security Intelligence, 243.

[78] Colaresi;, Democracy Declassified, 51.

[79] "MH17 - The Open Source Investigation, Three Years Later," Bellingcat, July 17, 2017, https://www.bellingcat.com/news/uk-and-europe/2017/07/17/mh17-open-source-investigation-three-years-later/.

regimes should be the most vulnerable to the effects of growing democratization of critical information. After all, such regimes withhold the most amount of information from public eye, have little state-society interaction in sharing political information, and have virtually no oversight against intelligence abuse. These regimes frequently spy on their own citizens with the explicit purpose of suppressing dissent and opposition, and due to the absence of safeguards, checks and balances, they suffer from structural mismanagement and corruption in national security and intelligence affairs. In contrast, although democracies will also suffer from drawbacks of exposure and leaks, such damage is thought to be minimal due to the existing democratic structures, including free and fair elections, a functioning parliament and public oversight and shaming mechanisms.

The biggest criticism of intelligence oversight and safeguard mechanisms is that they lack the technical knowledge and background to properly evaluate what their intelligence agencies are doing with technology.[80] This was best evidenced by some of the archaic and tone-deaf questions posed against Mark Zuckerberg during the Facebook testimony.[81] One major way OSINT can contribute to oversight is to provide readily-available analysis that most oversight mechanisms cannot conduct by themselves. Through a methodical analysis of open-source tools, a technically proficient networked crowd can aid more established, but slower safeguard institutions with data, evidence and monitoring metrics on the abuses of secrecy. But will OSINT expedite, or enable democratization of authoritarian regimes? This is unlikely, as there are more variables in this equation in real life. Although authoritarian states lose more substantial amounts of policy secrets to OSINT, this doesn't necessarily lead to a call to replace the regime, or government, or mobilize sufficiently to enable this transition. Most of the time, democratic leaks and exposures - as small as they can be - are more likely to lead in government resignations or substantial drop in government support, compared to autocracies. Regardless of their loss, authoritarian states are more likely to rely on brute-force tactics of imprisoning and intimidating potential blowback effects against exposure.[82] Although criticism and public reaction against exposures of mismanagement and miscalculations are similar across regime types, their ability to turn into political pressure and shake up a government are structurally different.

How about, then, the relationship between regime type and foreign policy effectiveness in the age of OSINT? The mainstream argument goes that the advent of OSINT makes it difficult for states to deceive the public or the international audience, given the availability of alternative information. Ideally, OSINT should enable a better flow of accurate information and proper fact-checking across the Internet, offsetting any propaganda effects of state-led misinformation attempts. This turn, is thought to make foreign policy more carefully-crafted and less likely to be based on deliberate misinformation, given their ultimate exposure through OSINT. However, this doesn't always turn out to be the case. One reason for this is the 'rallying effect'; an electoral reflex that translates into greater support and mobilization in support of the government and leadership during times of crisis and escalation. The rallying effect minimizes public reaction or resistance against lack of oversight and increases short-term tolerance against miscalculations.[83] This enables democracies and autocracies alike to make fast and potentially miscalculated decisions over the short-term; since most crises are inherently short-term, all regime types become more likely to make misjudgements, despite the fact that they operate in an OSINT-driven information environment. Evidenced by the empirical studies, authoritarian states too, suffer from audience costs in foreign policy, and democratic foreign policies are not necessarily more effective under information constraints compared to autocracies.[84]

[80] Amy B. Zegart, "The Domestic Politics of Irrational Intelligence Oversight," Political Science Quarterly 126, no. 1 (March 1, 2011): 1–25, https://doi.org/10.1002/j.1538-165X.2011.tb00692.x.

[81] Emily Stewart, "Lawmakers Seem Confused about What Facebook Does — and How to Fix It," Vox, April 10, 2018, https://www.vox.com/policy-and-politics/2018/4/10/17222062/mark-zuckerberg-testimony-graham-facebook-regulations.

[82] Michael M. Andregg and Peter Gill, "Comparing the Democratization of Intelligence," Intelligence and National Security 29, no. 4 (July 4, 2014): 487–97, https://doi.org/10.1080/02684527.2014.915174.

[83] John R. Oneal and Anna Lillian Bryan, "The Rally 'round the Flag Effect in U.S. Foreign Policy Crises, 1950–1985," Political Behavior 17, no. 4 (December 1, 1995): 379–401, https://doi.org/10.1007/BF01498516.

[84] Weeks, "Autocratic Audience Costs"; Branislav L. Slantchev, "Politicians, the Media, and Domestic Audience Costs," International Studies Quarterly 50, no. 2 (June 1, 2006): 445–77, https://doi.org/10.1111/j.1468-2478.2006.00409.x; Jessica Chen Weiss, "Authoritarian Signaling, Mass Audiences, and Nationalist Protest in China," International Organization 67, no. 1 (January 2013): 1–35, https://doi.org/10.1017/S0020818312000380.

Conclusion: Implications for International Security

The wider digital OSINT debate concerns how technology is changing the nature of state secrets and role of secrecy in statecraft. Until an equilibrium is established, communication technologies remain a battleground between states and their respective societies, as well as among states. As with past technological advances in communication – printing press, radio, television, satellites – Internet-based communication too, will enable significant social forces to push for greater liberties, and states, to repress such forces. From the state point of view, OSINT will lead to two outcomes. The short-term outcome will be a review of military and intelligence policies to prevent leaks and exposures through new communication tools. This will include simple behavioural adjustments, from smart phone use, to social media presence, including changing the way important political secrets are encrypted and stored. Over the long-term however, citizen-led crowdsourced OSINT initiatives will continue to expose government secrets and especially prevent states to dominate the narrative during crises and emergencies. Democracies and authoritarian governments alike will try to assert their own version of events, but will find it increasingly hard to establish a monopoly over the framing and narrative of important events. This will force governments either to suppress and block public mechanisms of alternative information, or change the way they utilize secrecy in statecraft. One example is how the press, internal leaks and public pressure combined have forced Bush-era detention facilities to be closed down under the Obama administration, resulting in the 2015 outlawing of the US Congress of all such facilities. However, similar pressures over the exposure of the Russian downing of the MAS17 airline in 2014 hasn't changed Russian behaviour – with the exception of restricting soldiers' cell phone use in combat zones. Similarly, the exposure of Russian soldier selfies in Crimea in 2014 had no effect on Russia's wider ambitions and operational course in Ukraine.

Therefore, it is unlikely that the advent of mass open-source analytics will have the same effect on all states. Nor is there evidence that OSINT will force all states to rely less on secrecy. Most likely, digital OSINT will create a 'secrecy asymmetry' between states – between those that have high tolerance to audience costs (i.e. autocracies) and those that are more responsive to them (democracies). Autocracies will find digital crowdsourced OSINT increasingly irrelevant in the wider scheme of things (except perhaps in critical operations) as leaks, exposure and citizen-led efforts can be offset through domestic tools of repression; arrest, jailing and censorship. Democracies on the other hand, will have to follow a different trajectory. This trajectory consists of alternative policy options that have to do with;

- Reforming public diplomacy agencies from a unidirectional posture (i.e. delivery of state position to the wider audience) to a multi-directional one, which involves disseminating public view and sentiment to government agencies, driving their adaptation to the digital open-source environment.
- Co-opting a degree of civilian crowdsourced OSINT into state intelligence efforts. This is less risky for more representative and freer political systems, where the amount of secrets that aren't already public knowledge is low. In contrast, this is hard for authoritarian governments that tend to be 'secrecy hoarders' and have much to lose (leak) through cooperating with public OSINT platforms
- Yield to greater judicial and legislative oversight in intelligence practice. By rendering intelligence operations more open to, and cooperative with safeguards, agencies can suffer less from audience costs in case some of their secrets are exposed through OSINT tools.

Over the long-term, the Internet and social media platforms will settle into a business equilibrium where the states will reassert their dominance over the flow of information, either through controlling the big technology companies, or reaching a power-sharing agreement that clearly defines jurisdictional areas to minimize leaks and exposures. Until then, such leaks and exposures will continue and will render states at a disadvantage against civilian-led analytics initiatives, and also create a new layer of security dilemma that will fuel international security competition and intelligence agencies' 'secrecy wars'. However, even in democracies, audience costs must not be exaggerated given the fact that online audience attention span is always limited and not directly linked to policy engagement. Social media engagement very rarely translates into actual political mobilization, and it is only when such social media engagement ends up creating a political, judicial or legislative momentum that OSINT efforts lead to real change. To that end, OSINT will increasingly find it more useful to pick its fights sparingly and focus its efforts on issues that are likely to generate wider public attention and policy momentum.

Ultimately, secrecy is not ending, but how we understand and think about it is rapidly changing due to open-information platforms. Events and facts that the states and societies used to think as secrets, are no longer secrets. This naturally brings about the necessity to rethink what to hide and what to disclose on the Internet, along with how to contain damage once these secrets are out. Until states and citizens adapt to new communication and information-extraction platforms, secrecy will remain a highly-blurred concept and will affect all sides of the state-society debate.

References

Aid, Matthew M. “All Glory Is Fleeting: Sigint and the Fight Against International Terrorism.” *Intelligence and National Security* 18, no. 4 (December 1, 2003): 72–120. https://doi.org/10.1080/02684520310001688880.

Andregg, Michael M., and Peter Gill. “Comparing the Democratization of Intelligence.” *Intelligence and National Security* 29, no. 4 (July 4, 2014): 487–97. https://doi.org/10.1080/02684527.2014.915174.

Appel, Edward J. *Cybervetting: Internet Searches for Vetting, Investigations, and Open-Source Intelligence, Second Edition*. CRC Press, 2014.

Apuzzo, Matt. “Who Will Become a Terrorist? Research Yields Few Clues.” *The New York Times*, December 21, 2017, sec. World. https://www.nytimes.com/2016/03/28/world/europe/mystery-about-who-will-become-a-terrorist-defies-clear-answers.html.

Asghar, Muhammad Zubair, Shakeel Ahmad, Afsana Marwat, and Fazal Masud Kundi. “Sentiment Analysis on YouTube: A Brief Survey.” *ArXiv* 1511.09142 (November 29, 2015). http://arxiv.org/abs/1511.09142.

Bacastow, Todd S., and Dennis Bellafiore. “Redefining Geospatial Intelligence.” *American Intelligence Journal* 27, no. 1 (2009): 38–40.

Baldino, Daniel, ed. *Democratic Oversight of Intelligence Services*. Sydney: Federation Press, 2010.

Born, Dr Hans, and Ms Marina Caparini. *Democratic Control of Intelligence Services: Containing Rogue Elephants*. Ashgate Publishing, Ltd., 2013.

Cadwalladr, Carole. “‘I Made Steve Bannon’s Psychological Warfare Tool’: Meet the Data War Whistleblower.” *The Guardian*, March 18, 2018, sec. News. http://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump.

Chang, Alvin. “The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram.” Vox, March 23, 2018. https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram.

Chen, Feng, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos, and Xiaohui Rong. “Data Mining for the Internet of Things: Literature Review and Challenges.” *International Journal of Distributed Sensor Networks* 11, no. 8 (August 18, 2015): 431047. https://doi.org/10.1155/2015/431047.

Chen, Hsinchun. *Dark Web: Exploring and Data Mining the Dark Side of the Web*. Integrated Series in Information Systems. New York: Springer-Verlag, 2012. //www.springer.com/gp/book/9781461415565.

Choi, Moonsun, Michael Glassman, and Dean Cristol. “What It Means to Be a Citizen in the Internet Age: Development of a Reliable and Valid Digital Citizenship Scale.” *Computers & Education* 107 (April 1, 2017): 100–112. https://doi.org/10.1016/j.compedu.2017.01.002.

Chulov, Martin. “Syria Attack: Nerve Agent Experts Race to Smuggle Bodies out of Douma.” *The Guardian*, April 12, 2018, sec. World news. http://www.theguardian.com/world/2018/apr/12/syria-attack-experts-check-signs-nerve-agent.

Colaresi;, Michael P. *Democracy Declassified: The Secrecy Dilemma in National Security*. Oxford, UK: Oxford University Press, 2014.

Collins, Dylan. “A US Airstrike Which Killed 38 People Allegedly Hit a Peaceful Mosque in a Syrian Village.” *Business Insider*, April 18, 2017. http://www.businessinsider.com/us-airstrike-allegedly-hit-a-peaceful-mosque-in-a-syrian-village-2017-4.

Davies, Philip H. J. “Intelligence Culture and Intelligence Failure in Britain and the United States.” *Cambridge Review of International Affairs* 17, no. 3 (October 1, 2004): 495–520. https://doi.org/10.1080/0955757042000298188.

De Stefano, Valerio. “The Rise of the Just-in-Time Workforce: On-Demand Work, Crowdwork, and Labor Protection in the Gig-Economy.” *Comparative Labor Law & Policy Journal* 37 (2016 2015): 471.

Desch, Michael C. “Democracy and Victory: Why Regime Type Hardly Matters.” *International Security* 27, no. 2 (October 1, 2002): 5–47. https://doi.org/10.1162/016228802760987815.

Dover, Robert, Michael S. Goodman, and Claudia Hillebrand, eds. *Routledge Companion to Intelligence Studies*. Routledge, 2013.

Dudczyk, J., J. Matuszewski, and M. Wnuk. “Applying the Radiated Emission to the Specific Emitter Identification.” In *15th International Conference on Microwaves, Radar and Wireless Communications (IEEE Cat. No.04EX824)*, 2:431–434 Vol.2, 2004. https://doi.org/10.1109/MIKON.2004.1357058.

Evans, Jacqueline R., Christian A. Meissner, Susan E. Brandon, Melissa B. Russano, and Steve M. Kleinman. “Criminal versus HUMINT Interrogations: The Importance of Psychological Science to Improving Interrogative Practice.” *The Journal of Psychiatry & Law* 38, no. 1–2 (March 1, 2010): 215–49. https://doi.org/10.1177/009318531003800110.

Fisher, Max. “Did Ukraine Rebels Take Credit for Downing MH17?” Vox.com, July 17, 2014. https://www.vox.com/2014/7/17/5913089/did-this-ukrainian-rebel-commander-take-credit-for-shooting-down-the.

Gettleman, Jeffrey. “Congo: Rapes by Civilians Rise Sharply, Study Says.” *The New York Times*, April 14, 2010, sec. Africa. https://www.nytimes.com/2010/04/15/world/africa/15briefs-congo.html.

Gibney, Mark. “The Downing of MH17: Russian Responsibility?” *Human Rights Law Review* 15, no. 1 (March 1, 2015): 169–78. https://doi.org/10.1093/hrlr/ngu036.

Giridharadas, Anand. “Ushahidi - Africa’s Gift to Silicon Valley: How to Track a Crisis.” *The New York Times*, March 13, 2010, sec. Week in Review. https://www.nytimes.com/2010/03/14/weekinreview/14giridharadas.html.

Glassman, Michael, and Min Ju Kang. “Intelligence in the Internet Age: The Emergence and Evolution of Open Source Intelligence (OSINT).” *Computers in Human Behavior* 28, no. 2 (March 1, 2012): 673–82. https://doi.org/10.1016/j.chb.2011.11.014.

Goldman, Zachary K., and Samuel J. Rascoff. *Global Intelligence Oversight: Governing Security in the Twenty-First Century*. Oxford: Oxford University Press, 2016.

Greenemeier, Larry. “DARPA Verigames Crowdsourced Formal Verification (CSFV) Project.” *Scientific American*, June 9, 2015. https://www.scientificamerican.com/citizen-science/darpa-verigames-crowdsourced-formal-verification-csfv-project/.

Gunitsky, Seva. “Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability.” *Perspectives on Politics* 13, no. 1 (March 2015): 42–54. https://doi.org/10.1017/S1537592714003120.

Gutierrez, Pablo, and Paul Torpey. “How Digital Detectives Say They Proved Ukraine Attacks Came from Russia.” *The Guardian*, February 17, 2015, sec. World news. http://www.theguardian.com/world/2015/feb/17/ukraine-russia-crossborder-attacks-satellite-evidence.

Harris, Mark. “How A Lone Hacker Shredded the Myth of Crowdsourcing.” WIRED, September 2, 2015. https://www.wired.com/2015/02/how-a-lone-hacker-shredded-the-myth-of-crowdsourcing/.

Hern, Alex. “Fitness Tracking App Strava Gives Away Location of Secret US Army Bases.” *The Guardian*, January 28, 2018, sec. Technology. http://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

Johnson, Loch K. *The Oxford Handbook of National Security Intelligence*. Oxford University Press, 2010.

Khalil, Osamah F. *America’s Dream Palace: Middle East Expertise and the Rise of the National Security State*. Cambridge, Massachusetts: Harvard University Press, 2016.

Klausen, Jytte. “Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq.” *Studies in Conflict & Terrorism* 38, no. 1 (January 2, 2015): 1–22. https://doi.org/10.1080/1057610X.2014.974948.

Landau, S. “Making Sense from Snowden: What’s Significant in the NSA Surveillance Revelations.” *IEEE Security Privacy* 11, no. 4 (July 2013): 54–63. https://doi.org/10.1109/MSP.2013.90.

Larkin, Sean P. “The Age of Transparency.” *Foreign Affairs*, April 18, 2016. https://www.foreignaffairs.com/articles/world/2016-04-18/age-transparency.

Legro, Jeffrey W. “Culture and Preferences in the International Cooperation Two-Step.” *American Political Science Review* 90, no. 1 (March 1996): 118–37. https://doi.org/10.2307/2082802.

Liptak, Andrew. “Strava’s Fitness Tracker Heat Map Reveals the Location of Military Bases.” *The Verge*, January 28, 2018. https://www.theverge.com/2018/1/28/16942626/strava-fitness-tracker-heat-map-military-base-internet-of-things-geolocation.

Lohr, Steve. “In Relief Work, Online Mapping Yet to Attain Full Potential.” *The New York Times*, March 28, 2011, sec. Business Day. https://www.nytimes.com/2011/03/28/business/28map.html.

Lyon, David, Kirstie Ball, and Kevin D. Haggerty. *Routledge Handbook of Surveillance Studies*. New York: Routledge, 2012.

Masciandaro, Donato. “Financial Supervisory Unification and Financial Intelligence Units.” *Journal of Money Laundering Control* 8, no. 4 (October 1, 2005): 354–70. https://doi.org/10.1108/13685200510620858.

McFate, Montgomery. “The Military Utility of Understanding Adversary Culture.” Arlington, VA: DTIC, Office of Naval Research, January 2005. http://www.dtic.mil/docs/citations/ADA479862.

McFate, Montgomery, and Steve Fondacaro. “Cultural Knowledge and Common Sense.” *Anthropology Today* 24, no. 1 (February 1, 2008): 27–27. https://doi.org/10.1111/j.1467-8322.2008.00562.x.

“MH17 - The Open Source Investigation, Three Years Later.” bellingcat, July 17, 2017. https://www.bellingcat.com/news/uk-and-europe/2017/07/17/mh17-open-source-investigation-three-years-later/.

Moore, Rowan. “Forensic Architecture: The Detail behind the Devilry.” *The Observer*, February 25, 2018, sec. Art and design. http://www.theguardian.com/artanddesign/2018/feb/25/forensic-architects-eyal-weizman.

Mueller, Hannes, and Christopher Rauh. “Reading Between the Lines: Prediction of Political Violence Using Newspaper Text.” *American Political Science Review*, December 2017, 1–18. https://doi.org/10.1017/S0003055417000570.

Oneal, John R., and Anna Lillian Bryan. “The Rally ’round the Flag Effect in U.S. Foreign Policy Crises, 1950–1985.” *Political Behavior* 17, no. 4 (December 1, 1995): 379–401. https://doi.org/10.1007/BF01498516.

Pringle, Robert W. “The Limits of OSINT: Diagnosing the Soviet Media, 1985-1989.” *International Journal of Intelligence and CounterIntelligence* 16, no. 2 (April 1, 2003): 280–89. https://doi.org/10.1080/08850600390198706.

Reiter, Dan, Allan C. Stam, and Alexander B. Downes. “Another Skirmish in the Battle over Democracies and War.” *International Security* 34, no. 2 (September 30, 2009): 194–204. https://doi.org/10.1162/isec.2009.34.2.194.

Richelson, Jeffrey T. “MASINT: The New Kid in Town.” *International Journal of Intelligence and CounterIntelligence* 14, no. 2 (April 1, 2001): 149–92. https://doi.org/10.1080/088506001300063136.

Rudner, Martin. “Britain Betwixt and Between: Uk SIGINT Alliance Strategy’s Transatlantic and European Connections.” *Intelligence and National Security* 19, no. 4 (December 1, 2004): 571–609. https://doi.org/10.1080/0268452042000327528.

Sanchez, Andy. “Leveraging Geospatial Intelligence (GEOINT) in Mission Command.” Arlington, VA: DTIC, Office of Naval Research, March 21, 2009. http://www.dtic.mil/docs/citations/ADA506270.

Schultz, Kenneth A. “Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War.” *International Organization* 53, no. 2 (ed 1999): 233–66. https://doi.org/10.1162/002081899550878.

Singh, V. K., D. Mahata, and R. Adhikari. “Mining the Blogosphere from a Socio-Political Perspective.” In *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, 365–70, 2010. https://doi.org/10.1109/CISIM.2010.5643634.

Slantchev, Branislav L. “Politicians, the Media, and Domestic Audience Costs.” *International Studies Quarterly* 50, no. 2 (June 1, 2006): 445–77. https://doi.org/10.1111/j.1468-2478.2006.00409.x.

Souza, Renato Rocha, Flavio Codeco Coelho, Rohan Shah, and Matthew Connelly. “Using Artificial Intelligence to Identify State Secrets.” *ArXiv* 1611.00356 (November 1, 2016). http://arxiv.org/abs/1611.00356.

Stewart, Emily. “Lawmakers Seem Confused about What Facebook Does — and How to Fix It.” *Vox*, April 10, 2018. https://www.vox.com/policy-and-politics/2018/4/10/17222062/mark-zuckerberg-testimony-graham-facebook-regulations.

Sullivan, Ben. “The Islamic State Conducted Hundreds of Drone Strikes in Less Than a Month.” *Motherboard*, February 21, 2017. https://motherboard.vice.com/en_us/article/vvxbp9/the-islamic-state-conducted-hundreds-of-drone-strikes-in-less-than-a-month.

Sutter, John D. “Ushahidi: How to ‘crowdmap’ a Disaster.” *CNN Labs*, October 25, 2010. http://www.cnn.com/2010/TECH/innovation/10/25/crowdmap.disaster.internet/index.html.

Thony, John Frank. “Processing Financial Information in Money Laundering Matters: The Financial Intelligence Units.” *European Journal of Crime, Criminal Law and Criminal Justice* 4 (1996): 257.

Tongur, Stefan, and Mats Engwall. “The Business Model Dilemma of Technology Shifts.” *Technovation* 34, no. 9 (September 1, 2014): 525–35. https://doi.org/10.1016/j.technovation.2014.02.006.

Traynor, Ian. “Libya: Nato Bombing of Gaddafi Forces ‘Relying on Information from Rebels.’” *The Guardian*, May 18, 2011, sec. World news. http://www.theguardian.com/world/2011/may/18/libya-nato-bombing-benghazi-rebel-leaders.

Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven ; London: Yale University Press, 2017.

Tzu, Sun. *The Art Of War*. Sterling Publishers Pvt. Ltd, 2005.

Weede, Erich. “Democracy and War Involvement.” *Journal of Conflict Resolution* 28, no. 4 (December 1, 1984): 649–64. https://doi.org/10.1177/0022002784028004004.

Weeks, Jessica L. “Autocratic Audience Costs: Regime Type and Signaling Resolve.” *International Organization* 62, no. 1 (January 2008): 35–64. https://doi.org/10.1017/S0020818308080028.

Weiss, Jessica Chen. “Authoritarian Signaling, Mass Audiences, and Nationalist Protest in China.” *International Organization* 67, no. 1 (January 2013): 1–35. https://doi.org/10.1017/S0020818312000380.

Westin Alan F. “Social and Political Dimensions of Privacy.” *Journal of Social Issues* 59, no. 2 (April 29, 2003): 431–53. https://doi.org/10.1111/1540-4560.00072.

Wheaton, Sarah. “New Technology Generates Database on Spill Damage.” *The New York Times*, May 4, 2010, sec. U.S. https://www.nytimes.com/2010/05/05/us/05brigade.html.

Wigell, Mikael. “Mapping ‘Hybrid Regimes’: Regime Types and Concepts in Comparative Politics.” *Democratization* 15, no. 2 (April 1, 2008): 230–50. https://doi.org/10.1080/13510340701846319.

Zegart, Amy B. “The Domestic Politics of Irrational Intelligence Oversight.” *Political Science Quarterly* 126, no. 1 (March 1, 2011): 1–25. https://doi.org/10.1002/j.1538-165X.2011.tb00692.x.

Zeitzoff, Thomas. “How Social Media Is Changing Conflict.” *Journal of Conflict Resolution* 61, no. 9 (October 1, 2017): 1970–91. https://doi.org/10.1177/0022002717721392.

Centre for Economics
and Foreign Policy Studies

Cyber Governance and Digital Democracy 2018/8

July 2018

Digital Open Source Intelligence and International Security: A Primer

H. Akın Ünver | EDAM, Oxford CTGA & Kadir Has University