Using Machine Learning to Understand the National Security Agency's Data Surveillance Trends

by

Elliot Akwanfo Damasah

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Digital Humanities

University of Alberta

Abstract

For over six decades, the public has remained ignorant about the National Security Agency (NSA) and its activities and has been shielded from the agency's invasive and unlawful projects. While the NSA's activities have proven valuable to the United States and its allies, it sometimes undertakes projects that contravene human rights. Despite the power it possesses, knowledge about the NSA has remained scarcely accessible to the public. But the dynamics changed after Edward Snowden divulged and distributed NSA data to some journalists, giving us opportunities to explore the agency to increase our understanding of their work. As such, this study aims to add to the literature about the NSA by ascertaining the NSA's discussions in SIDtoday, an online-based internal newsletter. It also aims to determine the trends of the various discussions within the agency and corroborate the agency's discussions of data surveillance (an intentional observation of people as a means for collecting data).

To test the hypothesis that the NSA discussed data surveillance issues, I collected text from the NSA's SIDtoday newsletters and analyzed them with a machine learning algorithm called topic modeling. Topic modeling is a text analysis method capable of revealing themes within a large collection of text. The purpose was to identify discussions and trends in the collection of text. The results revealed that the NSA discussed eighteen (18) main topic areas, and included in these topics were Data Gathering and Target Analysis. Further analysis of these two topics indicated that the NSA engaged in data surveillance. Additionally, the trend analysis revealed the agency's sustained and gradual increase in data surveillance from 2003 to 2006.

From the results, it became evident that the NSA engaged in data surveillance. The other topics broached by the algorithm revealed the banal activities of the agency and its labor force. The trend analysis results affirm the agency's inclination towards increased surveillance. Based on these findings, the public should be aware that the NSA is relentless in its efforts to undertake its surveillance responsibilities. That said, some people, especially activists and investigative journalists, ought to be vigilant in matters relating to their privacy.

Preface

This thesis is an original work by Elliot Akwanfo Damasah. No part of this thesis has been previously published.

Acknowledgements

From the beginning of this thesis, I have received several supports and assistance.

First, I would like to thank God for being with me throughout this journey. I would also like to thank my supervisor, Professor Geoffrey Rockwell, for his guidance, insightful feedback, and introduction to valuable resources that helped me to complete this thesis. Also, I am thankful to him for helping me formulate the research questions and methodology. Despite his busy schedule, he still managed to offer me tailored and timely advice.

To Dr. Harvey Quamen, thank you for being part of this thesis and for the constructive feedback you provided. Your input, in many ways, contributed to improving this thesis. My sincere thanks to Dr. Ali Shiri for investing his time and effort into making this thesis a success.

I cannot overlook the help of Nicola DiNicola and David Sulz. To all of you, I say thank you for your inputs.

In addition, I would like to thank my parents – Mr. Julius Damasah and Elizabeth Laryea – for their unwavering support throughout my life. Indeed, I could not have come this far without their assistance. To my friends -- Dr. Dominic Roberts, Dr. Emmanuel Kyeremeh, and Philip Akude -- thank you for proofreading this thesis and for providing stimulating discussions that helped me to cope with its demands.

Table of Contents

Chapter One: Introduction and Related Works *1*

Introduction..1

National Security Agency (NSA) 4

Birth of the National Security Agency..15

Edward Snowden and the Leaked Documents 21

Life of Edward Snowden, the Whistleblower 21

NSA Files and How they Came into Our Hands...31

Study Rational 43

Chapter Two: Methodology *45*

Data Preparation 45

a. Description of the data (SIDtoday)..45

b. Data Gathering...51

c. Extraction and Preprocessing 53

Analysis 58

a. What are the common themes and issues that come up in the NSA's newsletter (SIDtoday)?........58

b. Does any of the topics in SIDtoday relate to data surveillance? What are their labels? 68

c. What trends do we notice about NSA's data surveillance topics?..69

Chapter 3: Results and Discussion.. *72*

Topics Discussed by the NSA in SIDtoday 73

Topics that are related to Data Surveillance 86

Dynamics in Data Surveillance 89

Topic 12 89

Topic 18 (Data Collection)..98

Limitations 106

Chapter 4: Conclusion ***109***

List of References ***113***

Appendix A ***128***

List of Tables

Table 1. SIDtoday newsletters according to publication years. 53

Table 2. Topics with their Labels, Representation, and Term 73

Table 3. Frequency Count of these Terms: “new,” “tool,” and “tools.” 94

List of Figures

Figure 1. 48

Figure 2. 62

Figure 3. 65

Figure 4. 1 75

Figure 4. 2 76

Figure 4. 3. 76

Figure 4. 4. 78

Figure 4. 5. 78

Figure 4. 6. 79

Figure 4. 7. 80

Figure 4. 8. 81

Figure 4. 9. 81

Figure 4. 10. 81

Figure 4. 11. 82

Figure 4. 12. 82

Figure 4. 13. 83

Figure 4. 14. 84

Figure 4. 15. 84

Figure 4. 16. 85

Figure 4. 17. 85

Figure 4. 18. 85

Figure 5. 1 90

Figure 5.1. 1 91

Figure 6. 1. 93

Figure 6. 2. 96

Figure 7. 1. 99

Figure 7. 2. 103

Figure 7. 3. 104

Chapter One: Introduction and Related Works

Introduction

Before Edward Snowden's whistleblowing episode in 2013, the general public knew little about the National Security Agency (NSA) and its spying activities. Even for historians and experts in cryptography, spying, and cybersecurity, most information about this agency had remained inaccessible. Among authors who have focused on bringing public knowledge about the NSA and its works, only a minute few, including James Bamford, have had some success (Bamford, 2005). Yet Bamford and his colleagues struggle to unravel many facts and activities about the agency, proving that the agency's concealment efforts have been successful. Other authors, including Aid (2001), have also not been able to transcend the impedance to knowing more about the NSA, proving our limited knowledge about the agency.

Despite several previous attempts by many scholars to know what happens within the NSA, our knowledge has remained limited until recently. This inadequate knowledge in our possession is primarily because of the NSA's own inclination to keep its activities hidden. This responsibility of safeguarding information about the agency does not only operate at the institutional level. It also happens at the managerial and other leadership levels. For example, NSA's first director – Lieutenant General Ralph J. Canine – made it a priority to conceal, as much as possible, any information about their operations from the public. In staying true to the notion of being clandestine, Canine did not only hide from his office staff during his four-year term of office but also obscured information from Congressmen who asked him about the agency's activities (Bamford, 2018). To Canine's co-workers, his favorite response to Congressmen who inquired about the dealings of the NSA was this -- "Congressman, you don't really want to know the answer to that. You wouldn't be able to sleep at night." (Bamford, 2007, p.585). Undoubtedly, since its inception, the NSA has put much effort into preventing the public from knowing much about its operations.

Regardless of the NSA's concealment efforts, the narrative about its secrets decidedly changed when Snowden released classified documents about their activities to some journalists (Epstein, 2017). This release by Snowden allowed the public to learn more about the agency and its activities. Thus, because of the efforts of Snowden, the whistleblower, we are now aware that there exists an agency in the US that stops at nothing, including court restrictions, to invade people's privacy.

Not only do we know about its existence, but we also have the opportunity to learn more about this agency. Therefore, in this study, I analyze NSA's newsletters using computer-assisted text data methodologies to examine the historical activities of the agency. Thus, I ask the following questions:

a. What common topics or themes come up in the NSA's newsletters?
b. Do any of the topics in NSA's newsletter (SIDtoday) relate to data surveillance? What are their labels?
c. What trends do we notice about NSA's data surveillance topics?

Providing answers to these questions and adding background information would help to increase our knowledge about the NSA, further bridging the literature gap. In particular, the inclusion of the NSA's birth, Snowden's background, and the leaked documents help to build an understanding of the NSA's motivations for their surveillance activities and journey in the cryptologic world. Also, it helps to position the dataset in the context that explains the importance of studying it.

Obviously, through the leaked NSA documents, which are mainly textual, many opportunities abound not only to historians, cryptographers, and others interested in intelligence but also to digital humanists and data scientists who wish to research the agency using computer-assisted methodologies. Also, this inquiry about the NSA's activities will help us learn about some of the concerns that permeate within the agency. I believe the study will enable us to obtain insight into the agency's methods to surveil its targets.

Although understandable for the NSA to conceal information from the public, the lengths to which it went to illegitimately spy on the masses and hide their actions make it more important to analyze their documents. This analysis is possible through revealing scientific methods that will allow us to formulate and confirm hypotheses about their surveillance discussions. Additionally, the agency’s newsletter, SIDtoday, is a treasure trove of valuable information that could provide details of various past and ongoing projects within the agency. Therefore, the study seeks to take advantage of the opportunity available in SIDtoday to add to the current literature on the NSA’s culture and surveillance discourse in general.

To undertake studies on documents like SIDtoday, some scholars, including Rockwell and Sinclair (2016), have utilized computer-assisted text data analysis methods to reveal information that would otherwise not be possible with conventional analytical methods. In their book titled *Hermeneutica,* the duo used these computer analytic methods to analyze text data such as those in the Humanist discussions. Here, among the hypotheses they tested, they revealed that the change of name from Humanities Computing to Digital Humanities (DH) did not only mark an “administrative change” (p.75), but it also signaled a change in the way people used electronic text. Other scholars employed computer algorithms to study data to reveal latent and manifesto themes in textual data (Lee, Kim, & Yu, 2001). These analyses and their revelation depict DH's longstanding interest in theorizing and critiquing contents, but in this case, with augmentation by computers (Rockwell & Sinclair 2016).

In addition to the achievements realized through its use by Rockwell and Sinclair (2016), computer-assisted text analysis can help us conduct studies such as topic modeling, sentiment analysis, text categorization, information extraction, clustering, and visualization (Hotho, Nürnberger, & Paaß, 2005). Utilization of these text analysis approaches will not only make our work faster, freeing us from the laborious task of manually coding large volumes of text but will also help us realize consistencies that elude us when human coders take up the same responsibilities (Hotho, Nürnberger, & Paaß, 2005). Because of the advantage computer-assisted text analysis offers, I will apply it in this study to reveal the main issues of discussion within SIDtoday.

The first chapter provides an overview of the NSA, its birth, and how SIDtoday came into our hands. In the second chapter, I present a description of the data (SIDtoday newsletters) for this study and provide details about the methodology used to address the questions. Following the second chapter, I will present my results and discussions in the third chapter. The fourth chapter will contain my conclusion and recommendation for further studies, where necessary.

National Security Agency (NSA)

As indicated by the name, the NSA is the national level agency of the United States government responsible for the overarching tasks of meeting specific intelligence needs of the various security agencies in the country (Aid, 2001; Bamford, 2001). Aid (2001) and Bamford

(2001) describe it as the largest, perhaps, most powerful intelligence organization on earth. Its role is to provide for the US government and its various units products such as Signals Intelligence (SIGINT), a product obtained through a covert means of gathering intelligence from electronic signals and systems of another country, mainly adversaries (Howe, 1974; Aid, 2001).

The description of the agency goes further. For instance, on its website, the NSA identifies itself as the agency that "leads the U.S. Government in cryptology that encompasses both signals intelligence (SIGINT) and information assurance (now referred to as cybersecurity) products and services and enables computer network operations (CNO) to gain a decision advantage for the Nation and our allies under all circumstances." (NSA/CSS, n.d., para. 1). Also, Bamford (1982) reveals that the NSA is a national level agency of the United States mandated by the President to meet the military forces intelligence needs and the civilian security agencies. In his view, the agency is the central piece that handles most of the technical capabilities required for achieving the set intelligence goals for the various U.S. security agencies. McAvoy (2010) refers to the NSA as an agency that was established to "infiltrate the institutions of those countries solely for non-disruptive information gathering" and to "develop a unique and clandestine way to bring information back from all parts of the world" (p. 60). That said, the NSA is an agency that is heavily involved in data espionage through signal and electronic means and does so with an elaborate zeal for achieving its goals.

Collecting intelligence is not the only business of the NSA. The NSA is also responsible for Communication Security (COMSEC), the act of securing communication and communication systems of the US government from prying eyes (Aid, 2001; Bamford, 2001; Burn, 2005). Thus, COMSEC concerns all NSA activities that provide security for information communication among US government agencies (Aid, 2001; Howe, 1974). Through COMSEC, the US, while ensuring secure transmission of content to its recipients, mitigates or prevents unauthorized access to the said content, thereby guaranteeing authenticity, accountability, and authentication, among other things such as reliability.

SIGINT, the commodity of interest to the NSA, in addition to COMSEC, is comprised of other data gathering domains, including Communication Intelligence (COMINT) and Electronic Communication (ELINT) (Burns, 2005; Fitsanakis, 2007; Bernard, 2009). COMINT refers to intercepted signals from satellites, radios, microwaves, cellular, and now, fiber optic cables (Bamford, 2005). To take advantage of COMINT, the NSA positions itself at vantage points, enabling it to tap into the various communication channels and capture data, which they later process and analyze (Bamford, 2005). Like COMINT, ELINT, according to Wiley (2006), "is the result of observing the signals transmitted by radar systems to obtain information about their capabilities." (p. 1). Using ELINT, the NSA and its predecessors managed to acquire valuable information about their enemy's weapons, helping them to make informed strategic decisions (Aid, 2001). For instance, the NSA used ELINT to determine the capacity and power of the Protivovozdushnaya Oborona Strany (PVO), the Soviet's air command defense system, providing the US with priceless planning information (Aid, & Wiebes, 2001). Thus, SIGINT is composed of COMINT and ELINT, and each component focuses on a peculiar mode of gathering information to give the NSA an urge in its spying endeavors.

In terms of usage, the NSA is mostly not the ultimate user of its products as it has a laundry list of customers it distributes them to for decision making. Included on the NSA's list of customers are the National Security Council, the Secretary of State, the White House, the Secretary of Defense, the Secretary of Treasury, the Secretary of Energy, and the Secretary of Commerce (Horgan, 1991). It also includes the Central Intelligence Agency (CIA), Federal Bureau of

Investigation (FBI), Defense Intelligence Agency (DIA), the Joint Chiefs of Staff, military operational commanders, ally intelligence agencies, the three service intelligence agencies, and the military commands (Horgan, 1991). The NSA is responsible for meeting the intelligence needs of many top priority departments and agencies in its home country and abroad.

Financially, the NSA receives enormous support from the US government. During the post-second world war period in 1960, the NSA received funding for its 480-million-dollar expenditure, despite a White House recommendation against it (Loeb, 2001). In 1961, for instance, the agency, according to Bamford (1982), received 116.2 million, allocating 34.9 million dollars to research and development. One of the NSA's former Directors, William O. Studeman, despite the secretive nature of the agency, addressed a group to hint them about the agency, and in his speech relayed to his listeners that his agency received substantial financial support (Bamford, 2002). He, however, never mentioned if this budgetary dominance was only within the intelligence circles or all organizations in the US at large.

In another account by Bamford (2002), the agency is described as the biggest of its kind in the US considering the billions of dollars it receives for its expenditure, dwarfing the Central Intelligence Agency (CIA) and the Federal Bureau of Investigation (FBI) in that regard. Additionally, Bamford (2002) reveals that, at one time, the NSA received a large part of the Department of Defense's $1.4 billion budget out of a total of 2 billion dollars Congress allocated to all US intelligence activities. Joseph Goulden (1969) corroborates the budgetary facts presented by Bamford by asserting that the NSA's financial support was approximately $2 billion in 1969. As if the agency's budget was not big enough, between 2000 and 2001, Bamford (2002) mentions that its combined budget for the period was 7.3 billion. Undoubtedly, the NSA had a big budget, and its government was willing and able to meet its financial demands.

In terms of personnel, the NSA and other US SIGINT units have had many employees. For the numbers, by 1952, the NSA was comprised of more than 30,000 workers, consisting of both civilians and military men (Aid, 2001). The NSA's number shot up to 65,000 by 1960, which was more than twice their number eight years ago (Aid, 2001). Bamford (1983) and Johnson and Hatch (1998) contradict Aid's (2001) employee numbers by stating 10,000 as the figure the agency managed to retain in 1960, without any details on what month they recorded those figures. Similarly, in 1969, its employee numbers rose to about 95,000, with members of other intelligence agencies, and included in the 95,000 personnel were 15,000 – 16,000 based at NSA's headquarters (Aid, 2001). Fitsanakis (2007) confirms the personnel figures of Aid by stating that in 1969, the NSA's employees numbered close to 100,000. Although without providing a specific date, Bamford (1982) reveals the NSA had 38,000 more employees than the Federal Bureau of Investigation (FBI) and the Central Intelligence Agency (CIA). The staggering numbers, indeed, make NSA's labor force capability evident to anyone.

Like other agencies, the NSA falls under certain jurisdictions and regulations, but these controls were not in place from the beginning. According to Bamford (1982) and McAvoy (2010), in terms of boundaries and restrictions, the agency is under the authority of the Foreign Intelligence Surveillance Court (FISC), the entity responsible for implementing the Foreign Intelligence Surveillance Act (FISA) of 1978. Also, FISC was in charge of ensuring an oversight for agencies like the NSA regarding whose or what SIGINT they could obtain (Banks, 2006). The FISC was not in place until 1978, as the US Congress did not know about the NSA, and therefore, did not have any control over them (Bamford, 1982). As a result of a regulatory absence, the NSA took matters into its hands to conduct mass surveillance operations such as SHAMROCK and MINARET. However, the NSA terminated them after 23 years of operation (Bamford, 1982).

Because of the urgent circumstances under which president Truman ordered the NSA's creation and the US Congress's ignorance of its existence, the agency operated without supervision for a long time (Burns, 2005). The agency abused its power on surveillance before the FISC's establishment in 1978 (Burns, 2005; Vladeck, 2015). After establishing FISC, structure and order became prevalent to the agency despite their absence from the beginning.

Although the FISC's presence brought some degree of order in how the NSA went about its surveillance business, the court could not effectively control the agency's activities. Even after FISC, the agency continued to engage in illegal interception (McAvoy, 2010; Vladeck, 2015). Scholars have criticized FISC in many ways such as (1) the biased conduct of its proceedings, which in most cases went in favor of the NSA, approving the agency to undertake surveillance activities of any kind; and (2) the eccentric manner in which the Chief Justice of the US was the sole person to appoint judges for FISC (Vladeck, 2015). Not only do these facts prove the unquestioned support received by the agency, but it also confirms the negligence of the court that was to preside over its activities by scrutinizing cases brought to it for approval.

Because the NSA is a civilian agency responsible for meeting the intelligence needs of many departments and agencies in the US, particularly the military, it developed a mechanism for coordinating the distribution of intelligence resources (Burn, 2005). To achieve this distribution goal, the NSA created the Central Security Services (CSS). The CSS, established in 1972, is an agency that aligns with the NSA and is responsible for coordination between the US Armed Forces (military) and the NSA in matters concerning the provision of cryptologic support, knowledge, and assistance (Bamford, 2018; NSA/CSS, n.d.). Here, the NSA's role is to provide intelligence while the CSS handles the cooperation between the NSA, military, and civilian agencies (Bamford, 2018; NSA/CSS, n.d.). Thus, the NSA, in addition to providing intelligence to the agencies that rely on it, works with the CSS to ensure all interested parties are providing and receiving their intelligence request and needs respectively and unifying the cryptologic works of the Department of Defense (DoD) (Bamford, 2018). In this regard, the CSS plays the role of a middleman in this relationship. Also, for efficient workflow, the NSA director doubles as the head of the CSS to manage affairs at both ends.

In desperate times, the NSA adjusts its approach, and perhaps, allies with the United Kingdom, to meet the SIGINT needs of its customers. Such an adjustment became evident when Pretty Good Privacy (PGP), a free encryption program modeled on Marty Hellman and Whitfield Diffie's cryptologic system (McAvoy, 2010), stymied the NSA's dominant intelligence-gathering mode – SIGINT. The agency adopted its operations by allying with Human Intelligence (HUMINT) agencies, an intelligence-gathering approach through field agents (Aid, 2001). An alliance of this kind was once created between the NSA and the CIA, forming the Special Collection Services (SCS) (Aid, 2001; Bamford, 2001; Fitsanakis, 2007). The SCS's aim, according to Fitsanakis, was to obtain intelligence before it got encrypted, as the NSA was struggling in its decrypting endeavors. With this approach, the SCS resorted to means "ranging from computer hacking (HACKINT) and parabolic and tempest surveillance to recruiting foreign diplomatic staff with access to desired information" (Fitsanakis, 2007, p.552). Such alliances confirm the NSA's dedication to meeting the intelligence needs of its customers, regardless of who its partnered.

Further to NSA's modus operandi, the agency worked by secretly spying on other electronic communication mediums of its targets. In addition to accomplishing its goals by eavesdropping through mediums such as radio, satellite, microwave, cellular, and fiber optics, the NSA explores other intelligence-gathering means, including making deals with telecommunication companies that have access to data of persons for whom they provide services (Bamford, 2001; Greenwald & MacAskill, 2013). One such deal was PRISM, a program that tapped into the servers of major technology companies in the US, including Apple, AOL, Facebook, Google, PalTalk, Skype, and Yahoo (Greenwald, 2012; MacAskill & Greenwald, 2013). Through this program, the agency collected content such as search history, email contents, document downloads, and live chart contents (Greenwald & MacAskill, 2013; Gurnow, 2014). Before PRISM, the NSA ran a similar program called Stellar Wind under the Bush administration that, from 2001 to 2011, collected email and phone metadata (Greenwald & Ackerman, 2013). According to Greenwald & Ackerman (2013), the NSA then analyzes the collected data to glean intelligence. NSA's operations from these examples confirm the agency's willingness to obtain and process information, using any means to gain intelligence.

Also, in staying true to its objective of giving the U.S advantage in its decision-making processes through whatever means necessary, the agency has unlawfully engaged in projects such as SHAMROCK, MINARET, and others that are equally inappropriate to achieve this goal (Greenwald, 2012; Gellman & Poitras, 2013; Greenwald & MacAskill, 2013). Through alliances such as the Special Collection Services (SCS), it endeavored to circumvent encryption on the internet by taking advantage of Human Intelligence to bug offices of its targets (Brown, 2015). As a result of its actions, the agency received criticism (Espionage, 2013). In effect, the NSA utilized projects earmarked for mass surveillance to further its agenda.

In distress times, the NSA went to great lengths, including breaking the laws of the United States to achieve its goal of gathering relevant intelligence. According to Cauley (2006), one of such periods was after the September 11, 2011 bomb attack, when the agency expanded its surveillance operation to encompass all domestic users in the US. Although surveillance of this kind contravenes with the Fourth Amendment, the agency still proceeded without proper approval (Cauley, 2006; McInnis, 2009). In addition, Gellman and Poitras (2013) made known that with help from some of the major telecommunication companies in the US – AT&T, Verizon, and BellSouth, the agency managed to retrieve information, which they extracted and analyzed in the hope to detect any leads or traces of terrorism. As the agency tasked with gathering intelligence, in times tough times, the NSA did all in its power to provide for its customers their needed intelligence, even through lawless means.

Regardless, one thing that contributed to NSA's prestigious status in the intelligence community is the usefulness of the COMINT it provides its customers. Unlike other means for obtaining intelligence, NSA's COMINT approach had proven to be more useful and efficient, and the agency was willing to go to great lengths to get this resource for its customers (Burn, 2005). To corroborate the fact about COMINT, the Brownell panel, the committee ordered by President Truman to assess COMINT in the US, adjudged COMINT as the most impactful intelligence source for the US (Johnson, 1995). Without surprise, COMINT was referred to by a Top-Secret study in 1955 as the most accurate, timely, reliable, and thorough intelligence source for American agencies that needed it (Aid, 2001). As well, the former NSA Director, Bobby Ray Inman, in 1979, mentioned that NSA's SIGINT contribution brought immeasurable benefits to their national defense and foreign policies (Inman, 1979). Aid (2007) revealed that SIGINT had, by the end of the Cold War, with its impressive output, superseded the conventional espionage means -- spies, agents, defectors, diplomats, and Human Intelligence (HUMINT) -- of gathering intelligence. During President Dwight Eisenhower's time in office, Dr. George Kistiakowsky, a significant intelligence community figure, described the conventional intelligence gathering means as dreadful compared to COMINT (Burrows, 1999). Aid (2007), however, did not reveal to us the intelligence output from the other intelligence agencies for readers to assess their merits. Aid (2007) attributes COMINT's rise to the top to its achievements of producing results where other intelligence sources have failed. COMINT, NSA's main product, had proven extremely useful to the US, making the agency that gathers and analyzes them more reputable.

Further to NSA's acquired esteem in the US government and intelligence community, the agency made significant contributions that strengthened its stance throughout the years. In Aid's account, NSA's COMINT was the best intelligence source on the Soviet Union, providing details of their activities and capabilities. Some of the information COMNT helped reveal to the US intelligence community include Russia's lack of well-trained pilots, aerial refueling capabilities, and strategic bombers (Steury, 1996). Through COMINT, the US gathered information on the Soviet Union's transport and logistics, nuclear weapon tests, activities of the Soviets in Eastern Europe (Breckinridge, 1993). As if this information gathering ability was not impressive enough, COMINT enabled the US to gather intelligence on the Soviet Union on a minute-by-minute basis, granting the US the ability to ascertain their plans through anomalies from their routines. COMINT's benefit was not limited to military activities, nor did it benefit only the US, but it also helped America, Australia, and Britain to identify Soviet spies (Benson &Warner, 1996). The benefits of SIGINT were enormous to the US and its allies, making the NSA powerful in the intelligence community.

The agency remained faithful to one of its core values – to monitor everyone else, including its allies. Among countries that fell victims to this unruly espionage was Russia during WWII and Canada. In a paper published by Obar and Clement (2013), they stated that the agency intercepted internet communications packets of Canadian residents. Using the IXmaps software, the two managed to identify communication routes originating from Canada to be terminating in the U.S.

as their exchange points. Interestingly, the NSA has equipment at the various terminal locations that make copies of all packets that transit through them (Obar & Clement, 2013). To corroborate their findings, a former employee of AT&T divulged some secrets. Also, according to Klein and Bamford (2009), the NSA installed monitory devices at one of the telecommunication company's exchange stations in San Francisco that operated by duplicating all traffic packets that went through it for analysis. Like the agency has done in the past, it is not surprising that the NSA utilized whatever means to monitor whatever or whoever is of interest to them, even its allies.

Having appraised the NSA and its vital role in meeting the SIGINT Intelligence needs of the various U.S. intelligence organizations, we can conclude that the NSA is critical to its country. Since its establishment in 1952, the agency has collaborated with other U.S. security agencies and its ally countries to realize impactful SIGINT results, giving its consumers decisive advantages. Also, as the SIGINT generating agency for the most powerful country in the world that likewise appreciates the usefulness of intelligence, the NSA is unsurprisingly a core agency in the intelligence community. With its financial support, personnel, and executive support, it has been able to undertake projects that have been overwhelmingly beneficial to the U.S. and its allies (Howes, 1974; Burns, 2005).

Indeed, it is easy to draw mixed conclusions about the NSA. As the core national-level information agency specialized in espionage through electronic mediums and ensuring the safety of information systems of the US government, it has managed to safeguard the interest of the US and its allies. Undoubtedly, the NSA's prestigious status in the intelligence community is well-deserved, as it has supplied its various customers with reliable, timely, and accurate intelligence. And with its budget that sometimes runs into billions of dollars, it has been able to undertake huge, daring, and privacy-invasive projects that collect massive amounts of data about users, corporations, and countries all over the world, including internet users worldwide. To achieve its primary goal of providing intelligence to the U.S., the NSA relied not only on its capabilities but also on its allies. It did all in its power to carry out its projects, including contracting big technology corporations to assist its operations, mainly by obtaining data from them. The NSA is unrelenting, and in some cases, unscrupulous in its approaches to obtaining the needed information for its operations, allowing nothing to get in its way, not even the FISA court that is in place to regulate its operations. With its huge budget, the NSA has also managed to stay relevant to the U.S. by adjusting and resorting to new approaches to undertake its job. In worldwide surveillance, we can rank the NSA among the topmost echelons, if not the ultimate.

Birth of the National Security Agency

On 28 December 1951, the Brownell Committee made some impactful revelations about the challenges of the Armed Forces Security Agencies (AFSA) (Howe, 1974). Led by George Brownell, the committee, which came into existence after a directive from President Truman, through its inquiries, realized that the AFSA's goal of unifying COMINT activities in the US had been stymied by efforts of the Joint Chiefs of Staffs (JSC) (Burn, 2005). This revelation triggered General Walter Bedell Smith's DCI to draw the attention of the National Security Council (NSC) to the impediments of the AFSA, as well as the need for centralization of COMINT in the US (Burns, 2005). The committee consisted of non-military members selected primarily by the Central Intelligence Agency (CIA), a means of excluding the military agencies as they in the past worked against any effort to unify COMINT activities in the US environment (Howe, 1974; Aid, 2001; Burn, 2005; Fitsanakis, 2007).

From the Brownell Committee in June 1952 came a recommendation for a reorganization in the intelligence community and the establishment of a unified COMINT agency that would have total control over all COMINT operations and resources in the US. It also recommended that the new organization received a clearly stated mission with a written presidential backing that liberates it from the authority of the JCS and AFSAC (Armed Forces Security Agency Council) (Bamford, 2018). According to Bamford (2018), the report recommending that the AFSA be positioned directly under the Department of Defense stated that the new agency be also placed under the authority of the Secretary of Defense and the Secretary of State. Regarding policy matters, the report suggested that the new agency be controlled by the United States Communications Intelligence Board (USCIB), which should be chaired by the Director of Central Intelligence, the military units, and all the non-military agencies fairly represented manner (Burn, 2005).

Also in the report were some reviews on the existing structure in the Intelligence community. According to Burn (2005), the report revealed that the COMINT structures did not reflect a unified unit but rather an alliance between four (4) service units. Also, it highlighted the inadequate authorities received by the director of the AFSA, citing that this made it impossible for the director to control the service units (Burn, 2005). As a result, the USCIB, JCS, and AFSAC received criticisms for their negligence in assisting the AFSA to execute its job (Benson, 1997). Regardless, the report mainly focused on COMINT production, centralizing its authority, management, and policy oversight (Burn, 2005).

As far as the Brownell report went, President Truman and the NSC, in October 1952, considered the committee's findings, leading to a revision of the National Security Communications Intelligence Board (NSCIB No. 9) (Aid, 2001). Through the revised NSCIB No. 9, Truman emphasized that COMINT was a national asset and so introduced changes that prioritized national COMINT interest over those of the individual service units (Howe 1974; Aid, 2001; Fitsanakis, 2007). Judging from the recommendations and the approvals the report received, it was evident that the US authorities welcomed the idea of an agency that would control COMINT for the entire country.

The Brownell Committee's report had some recommendations for the reorganization and unification efforts of the COMINT structures. First, pushing against the military's sovereignty over COMINT, the Brownell report insisted that the intelligence resource is an asset that should receive national interest and attention, allowing the civilian organizations to share in this resource, according to Burn (2005). Second, although amidst some contemplations, the report recommended that control over COMINT be centralized (Benson, 1997). Focusing on the AFSA as the ideal agency for the centralization idea, the report outlined that the agency should unify and exercise control over COMINT activities in the US (Burn, 2005). Thirdly, the Brownell Committee also recommended that AFSA be accountable and responsible for its budgetary allocation, as its predecessors showed little evidence of accountability (Johnson, 1995).

After receiving the committee's report, Major General Ralph J. Canine added his comments and recommendations (Johnson, 1995). Arguing in favor of the majority principle, he stated in a letter that the report's suggestion to retain the unanimous rule principle would hinder the intelligence community's interest as the latter, in the past, negatively affected the work of the latter USCIB (Johnson, 1995). He also had comments about the report's assessment of COMSEC and NSCID No.9's application to the new directive. To Canine, the COMSEC aspect of the report did not include enough details, so he recommended that they make it the focus of another study (Benson, 1997). In addition, he commented that they apply to the new directive the strict rules of the USCIB (Benson, 1997). His review of the report identified key factors that helped to improve the overall effect of the report.

Relating to organizational structures, the report also suggested that AFSA's relationship to any authority above, within, and below it be reviewed. First, regarding powers above the AFSA (USCIB and the Department of Defense), in addition to calling for the termination of the AFSAC, the committee proposed that the agency be removed from the control of the JCS and be made a subordinate of the Department of Defense (Burn, 2005). For the powers within the AFSA, it proposed that AFSA's director be at least a three-star military officer and that his deputy must be a civilian (Benson, 1997). Likewise, he suggested that a civilian could become the director, but his deputy must be a military officer (Benson, 1997). Finally, for AFSA's relationship with powers below it, the report, as has mostly been the case, recommended that the AFSA, with a Presidential memorandum backing, be instituted as a principal agency for COMINT activities, as well as have the duty to unify intelligence activities for the federal government (Gilbert & Finnegan, 1993). Overall, through a brief communication, the report made recommendations about how the US authorities should handle the three levels of power structures relating to the AFSA.

Similar to the recommendations for the AFSA, the Brownell report had some suggestions for the USCIB. The report, apart from recommending a decrease in the military's representation on the USCIB, proposed that its board membership comprised of the Secretary of State, the Secretary of Defense, the Director of Central Intelligence (DCI), the Chairman of the Joint Intelligence Committee, the FBI, and the Director of AFSA (Burn, 2005). It further proposed that matters concerning the AFSA should be determined by a majority rule approach, in addition to the suggestion that the DCI becomes USCIB's chairman permanently (Burn, 2005). It recommended a greater responsibility for the USCIB over policy and harmonization of COMINT affairs (Howe, 1974). It is evident that the committee saw flaws in USCIB, triggering it to make changes.

After the series of reviews the Brownell report underwent, there came a time for implementation. On 24 October 1952, President Truman issued a memorandum for the establishment of the NSA and the restructuring of the USCIB to conform to the committee's recommendations (Burn, 2005). In addition to the Revised National Security Council Intelligence Directive Number 9 (NSCID No. 9), Truman, through the memorandum, which he implemented with help from the Secretary of Defense, Secretary of State, and the Department of State, caused the JCSs to lose their USCIB membership (Bamford, 2018). However, the director of the NSA received voting right; the three military services remained on the board; and the DCI became the permanent chair of the board (Benson, 1997). The new structure, further to ensure fair participation of the civilian and military agencies, put in place mechanisms for resolving issues related to intelligence in the US (Burn, 2005).

As the executive agent of the memorandum issued by Truman, Robert Lovett, Secretary of Defense, had to ensure the commencement of the NSA while addressing other intelligence issues such as COMSEC (Benson, 1997). According to Benson (1997), Lovett needed to address the COMSEC concerns because Truman's October 24, 1952 memorandum did not cover that. Although Truman, in a second memorandum, declared COMSEC as a national responsibility to be addressed by the USCIB, someone needed to be responsible for its implementation. Lovett assumed the responsibility and issued two memoranda to handle COMINT and COMSEC affairs (Johnson, 1995). Truman's attention to COMSEC indicated his value for keeping information confidential and from prying eyes.

On 4th November 1952, Lovett finally ensured the commencement of the NSA (Burn, 2005). Besides, he issued a memorandum to the Joint Chiefs of Staff (JCS) and NSA's director to inform them of the institutional changes that have taken place (Bamford, 1983). Among other pronouncements, Lovett stated that the Armed Force Security Agency (AFSA) had changed to the NSA and that all of its responsibilities and resources be allocated to the new agency. COMSEC responsibilities went to the NSA; Department of Defense’s COMINT resources were also to go under the NSA’s director’s command and the responsibility and task of formally implementing the NSCID No. 9 also went to the NSA’s director (Bamford, 1983; Aid, 2001). To lead the NSA, Lovett’s memorandum, on 4th November, appointed Major Canine as its director, and in compliance with Truman’s order, Canine received a third star (Burn, 2005). The agency, for the next four year of Canine’s term in office as NSA’s director, he had under its control all COMINT production and collection resources.

In addition to COMINT, COMSEC responsibilities later transferred to the NSA in 1952 and on condition that the NSA treated it like the AFSA did while awaiting the NSC’s recommended approach (Howe, 1974). Later in 1957, the NSC provided a permanent order that created a committee for COMSEC -- US COMSEC Board (USCSB) -- similar to COMINT's. The new order made room for COMSEC needs of both the military and non-military agencies to be met by means of including representatives of the various groups on the USCSB (Burn 2005). In the ensuing years, the Secretary of Defence delegated his executive powers to the Director of NSA, which empowered him to make decisions best for COMSEC matters. (Burn, 2005). Regarding those it served, the AFSA's COMSEC clients included NATO allies, U.S. National Military Establishment, the U.S (Howe, 1974).

The processes that led to the birth of the NSA were complicated, but the agency eventually became a reality. From the onset, the U.S Army and Navy initially took charge and monopolized issues about this vital resource, COMINT, but things changed when other parties developed an interest in it. The Army and Navy, eager to retain total control over COMINT did all in their powers to prevent other interested parties from getting involved. Also, to ensure that their individual interests were always a priority, they refused to work completely together on COMINT affairs. But this monopoly and lack of collaboration began to see opposition after the second world war, and even more so after the Pearl Harbor attack. President Truman, noticing the potential of SIGINT for the U.S., responded to calls from some military leaders who wanted to improve the SIGINT situation in the country. This response, which was also triggered by the hostility in the intelligence community, moved him to task the Brownell Committee to work on a solution that would benefit all involved parties. Although the Stone committee had done a similar job in the past, it could not bring the military COMINT units together, but the Brownell Committee made a difference. In 1952, the committee's report made recommendations that led to the transformation of the Stone recommended Armed Forces Security Agency to the National Security Agency, freeing it from the control of other authorities that suppressed it, leading to the birth of the NSA.

Edward Snowden and the Leaked Documents

Life of Edward Snowden -- the Whistleblower

Due to the efforts of Edward Snowden, we have access to more NSA documents than ever before, most of which are with journalists, including Glenn Greenwood. Regarding the number of documents, King (2014) reveals that about 1.7 million leaked top-secret documents of the

Nationals Security Agency were released by Snowden and are in the hands of some selected journalists.

Snowden, an IT professional who once worked for the CIA, Dell, and Booz Allen Hamilton as a contractor, was born in 1983 in Elizabeth City, North California (Epstein, 2017). In 1998, at age 15, he dropped out of Arundel High School, citing medical conditions as his reason for the decision (Gurnow 2014; Epstein, 2017). Gurnow (2014) reveals that Lon Snowden, the father of Edward Snowden, claimed that his son’s absence from school was due to mononucleosis, an infectious viral disease. Epstein believes this claim. However, Gurnow believed that Lon may have concealed the truth to the advantage of his son. Regardless, Snowden could not receive a High School Diploma (Epstein, 2017; Gurnow, 2014). Yet, he did not give up on his hope of attaining greater heights in life.

By 2001, his parents divorced, causing him to live by himself in Ellicott City, the city where his mother owned a house (Gurnow, 2014; Epstein, 2017). During this time, Epstein revealed that Snowden kept himself busy by posting about games on a website called Ars Technica. On Ars Technica, the account of Gurnow and Epstein suggested that Snowden showed an interest in playing video games and a desire to develop them. He might have picked up some useful computer skills during his time at home, which later played to his advantage. His 140 IQ, as stated by Gurnow, may have contributed to his ability to acquire the IT prowess he possessed. These IT skills, including those he picked up throughout his career, would later play vital roles in his plot to purloin NSA’s data. Nonetheless, his IT expertise brought him some benefits on his career path, as he secured a job as a webmaster with Ryuhana Press, a Japanese comic and animation business established by Snowden and his friends (Epstein, 2017; Gurnow, 2014).

In May 2004, Snowden enrolled in the US Army through 18X, a program for individuals who could not meet the required education criteria for enlisting into the Army (Gurnow, 2014; Epstein, 2017). However, he was relieved of his duty in September of that same year on administrative discharge, a discharge that results from undesirable conduct of service persons (Epstein, 2017). Despite staying unemployed for a couple of months, Snowden got a job as a security guard and then upgraded to a communication officer in 2006 with the CIA (Epstein, 2017). In an interview with The Guardian, Snowden stated that he left the Army because he broke his legs (Marbella et al., 2013). Gurnow asserts that Snowden might have gotten both of his legs broken from airborne training in the Army. Epstein, however, disagrees with the assertion that his legs were injured during training. He supports his claim by stating that none of Snowden's neighbors saw him with a walking aid. Possibly, Snowden's neighbors could not verify this information since he could have spent most of his time indoors because of his injury period.

In 2005, after his exit from the Army, Gurnow speculates that Snowden worked undercover at the University of Maryland Centre for Advanced Study of Language (CASL), but Epstein holds a contrary opinion. Drawing from Snowden's comments on Ars Technica, Epstein concludes that Snowden worked the night shift as a security guard during that time. Gurnow also pointed out that it was challenging to determine Snowden's exact position at CASL based on the confusing post on Ars Technica. Admittedly, it is somewhat difficult to arrive at the truth about Snowden's job position at the CASL given the varied opinions. His comments on Ars Technica makes it even more arduous. For instance, he wrote in one of his postings on Ars Technica that he "makes twice the average income (Gurnow 2006, p. 8)." While it is difficult to ascertain what he meant by "average income," it is even more strenuous to determine a job position based on an employee's income, as several job titles can have similar wages. Besides, Brian Ulmann, Spokesman for the

University of Maryland, stated that Snowden was a “security specialist” for CASL at the time (Finn, Miller, & Nakashima, 2013, para. 21). Because Snowden was in his early twenties at the time, it is unlikely that he would have acquired all the skills and certifications needed for a security specialist. Therefore, it is convincing enough to settle for the theory that Snowden worked as a security guard at CASL. Most importantly and convincing of all, Snowden said in an interview that he took the security guard job at CASL to gain a top security clearance at the institution, confirming that he did not hold a security specialist position at CASL (Greenwald, 2014).

Shortly after his employment at CASL, Snowden got an offer with the Central Intelligence Agency (CIA) in November 2006, which got him working at Geneva in 2007 (Gurnow, 2006; Epstein, 2017). Similar to his position at CASL, there were speculations about the exact role he occupied at this new job. Snowden also does not help matters. For instance, in a video interview on *The Guardian*'s website, Snowden told Greenwald that he was a “senior advisor” for the CIA (Greenwald & Poitras, 2013). According to the foreign affairs ministry of Switzerland, he was a “mission” employee, but rumors in the media speculated that he was conducting espionage for the US while using his network security job title as a cover (Ferran, 2013). The speculations invariably make it challenging to settle on any of the positions with a definite conclusion. But it is reasonable to conclude that he was in Geneva as a computer network security worker, as his IT skillset supports this conclusion adequately.

While in Geneva, Snowden began to assess the moral implications of the CIA's conduct in Switzerland, and his disapproval of their actions may have convinced him about his inclination to expose the US intelligence community member -- the NSA (Epstein, 2017). Mavenee Anderson, an intern who worked with Snowden in Geneva, later confirms that during Snowden’s tenure of employment with the CIA, some of the agency’s conducts troubled him (Epstein, 2017). For one, Snowden, in an interview with The Guardian, mentioned a CIA operation in Geneva in which the operatives encouraged a Swiss banker to drive while drunk to get him apprehended so that they could rescue him in return for a favor – obtain information in return (Greenwald, 2014). Snowden considered this action “unethical,” and according to Gurnow, this potentially spurred him into exposing the US government’s secret surveillance activities in 2009.

The CIA’s modus operandi may not have been the only factor that propelled him to plot against the Intelligence Community member. It is evident in Epstein’s writings that Snowden may have run into some troubles with his CIA superior, which stirred suspicion in his bosses. In a *New York Times* story, Eric Schmitt (2013) revealed that Snowden’s bosses suspected that he wanted to illegally get access to top-secret information. When the authorities got to know of this action, they investigated him, but they halted the process when Snowden resigned from the CIA, per an agreement between both parties (Schmitt, 2013). The resignation they forced on him could have been another contributing factor for the young man’s decision to work against the intelligence community's interest, including the NSA.

Further to the factors that started his troubles, Snowden claimed that in 2009, he drew the attention of his CIA superiors to the vulnerability in the computer system at his Switzerland station, but his bosses frowned on his actions (Epstein, 2017). He believed they could have mitigated the vulnerabilities if they had considered his prescribed precautions. To prove his point, he stated in an email to Risen (2013), a *New York Times* writer, that he made changes to software codes on the agency’s computer systems. His bosses, however, did not heed his suggestion, nor his unauthorized changes. Unfortunately, for Snowden, the CIA’s management did not like the idea of changing codes in their system. So, they reacted by stating that his attempt to access unauthorized files was in contempt to his agreement with the CIA, resulting in a “derogatory comment” on his evaluation file (Risen, 2013, para. 18). Consequently, the CIA forced him to resign, which he did in January 2009 (Epstein, 2017). In an interview with Risen, Snowden later claimed that a senior manager punished him for trying to reveal a vulnerability to the CIA. Amidst all the confusion, one thing is certain. This sequence of actions has the propensity to infuriate many people, and that the CIA should have anticipated his actions against the intelligence community in the US.

Despite deciding to expose government surveillance, Snowden claimed he refused to take action against the CIA at the time because he would have risked the lives of spies for the CIA had he done it (Greenwald & Poitras, 2013). While it is true that he did not reveal the identity of any CIA spy, it cannot be disputed that Snowden did not reveal any classified information about the agency because he may not have had his hands on any. The other reason he claimed convinced him to halt his decision to release the documents to the journalist was his expectation that President Barack Obama's administration would have implemented his desired changes – end government's invasive surveillance programs (Greenwald & Poitras, 2013). According to Greenwald (2014), his anticipation for Obama to make the needed reforms unfortunately never materialized but worsened with time. In an interview with Greenwald and Poitras, Snowden stated that "you can't wait around for someone else to act. I had been looking for leaders, but I realized that leadership is about being the first to act" (Greenwald & Poitras, 2013, para. 37). These sayings of Snowden expressed his frustration to do something about the surveillance situation in the US.

Snowden's opposition to government surveillance was not only expressed in his interaction with his colleagues, as it reflected in his political affiliations. According to Epstein (2017), he sympathized politically with "the right." In his book, *The Christian Right, the Far Right, and the Boundaries of American Conservatism,* Durham (2000) refers to "the right" as a group that holds conservative ideologies such as the rights to possess guns, fight against abortion, opposition to gay rights and feminism. In the US, these ideologies can be cursorily linked to the Republican party, implying that Snowden leaned in favor of the said party and their ideologies. His like of some political figures confirms political proclivity. Snowden was a supporter of Ron Paul, a libertarian political ideologist who held strong views against government surveillance (Gurnow, 2014; Epstein, 2017). To support this cause, Epstein suggests that Snowden even donated 500 USD to Ron's campaign (Blake, 2013). A donation of this kind shows how serious Snowden is about his views. It is, however, unclear whether Snowden became a supporter of Ron after he discovered government surveillance and labeling of the NSA's surveillance conducts as wrong. Regardless, Ron's view on government mass surveillance may have influenced Snowden's perspective.

Five months after leaving the CIA, Snowden got an offer at Dell, an IT company contracted by the NSA. Over his 45 months with Dell, he served in various IT capacities, including a cybersecurity trainer and a system administrator in Japan. Based on the account of Gurnow and Epstein, Snowden's new position gave him access to both NSA computers and encrypted NSA files. Although protected, Snowden again managed to poke around to find vulnerabilities in the system at NSA's facility, which he again brought to the attention of his superiors (Epstein, 2013). It appeared Snowden had a knack for identifying vulnerabilities. Nonetheless, they did nothing about the flaw Snowden reported, according to Epstein. The NSA may have ignored the vulnerability he reported because it was upgrading its systems at the time, and there was the possibility that the issue would resolve after the upgrade. Regardless of the flaws in the systems, Snowden was not at liberty to search for weaknesses in NSA's systems. If nothing at all, such actions made the senior management suspicious of him. Such inquisitiveness brought him nothing but unnecessary attention. Unlike at the CIA, he did not get into trouble for making his NSA bosses aware of the vulnerabilities in the system (Gurnow, 2014; Epstein, 2017).

In Epstein’s (2017) account, Snowden embarked on a trip to India in September 2009 for computer programing and hacking tutorials at Koening solutions. Koening is an IT training company that provides various computer certifications worldwide (Koenig, 1993). Epstein indicates that Snowden took a crash course in computer hacking, which provided him with the knowledge needed to detect and exploit security vulnerabilities at his post with the NSA. Snowden may have identified malicious opportunities in the NSA systems he worked on. Also, he may have realized that hacking skills were essential in his new objective of exposing US invasive surveillance. This exposé agenda of his could have been his primary reason for his trip to India.

At some point, Snowden became fully aware of the NSA’s conduct, compelling him to bring to the public’s knowledge the intelligence community's activities. As mentioned earlier, he found the agency’s surveillance program to be particularly worrisome, which he got to know of by illegally accessing the NSA’s inspector general’s report of 2009. To confirm his concerns, Mavenee, the intern who worked alongside Snowden, attested to his discomfort with the CIA’s surveillance programs and methods (Greenwald, 2014; Epstein, 2017). Regardless of the evidence that pointed at his tendency to bring harm to the agency, by March of 2012, Snowden was offered another position as a system admin at NSA’s Hawaii office, which he accepted (Gurnow, 2014; Epstein, 2017). It is quite ironic that someone who had issues with the conduct of the NSA would accept an offer by the same agency. There, however, was a catch for his acceptance of the offer. The new position allowed him to retrieve the documents needed to disclose the agency’s conduct to the public.

To increase his prospects of gaining access to more documents about the agency’s activities, as well as move up the ranks, Snowden took an NSA’s entrance exams with the hope of securing a high score. He assumed that a near-perfect score would get him a senior executive officer position (SES) (King, 2014). Employing his hacking skills, Snowden managed to get a hold of the exam questions, which he used to score a high mark, but unfortunately for him, he was denied the position he desired (King, 2014). This incident was the second time he used his hacking skills, according to Epstein (2017). Snowden had previously employed this skill at the CIA when he made an unsolicited change to some systems software codes. Epstein (2017) purported that Snowden wanted to gain access to files that were not accessible to him as a Dell employee, so he endeavored to gain employment as a senior executive at NSA to get the needed access. It became somewhat evident at this point that Snowden was plotting something devastating for the NSA.

Despite his failed efforts to rise to the level of a Senior Executive officer (SES), Snowden was still in an advantageous position to steal Top Secret NSA documents, as he was the System Administrator in charge of a backup transfer program from NSA's Fort Meade to their Hawaii facility (Epstein, 2017). Epstein reveals that Snowden's position made it easy to secretly copy data from the Hawaii facility, as they did not have any file monitoring system in place at the time. He took advantage of his position to advance his course. One such means was to use NSANet, a shared network platform where the NSA, CIA, and DIA (Defense Intelligence Agency) workers discussed their projects (Epstein, 2017). This platform provided Snowden with a massive ground access more information.

Confirming the magnitude of the data Snowden collected from NSANet, Michael Hayden, former NSA director, revealed to Epstein that Snowden copied valuable data from NSANet, which is composed of data that pertained not only to the NSA but also to the CIA and DIA. All these data theft activities were in preparation for his big revelation to the public.

Snowden later delved deeper in pursuit of more data to purloin. As of 2012, Epstein suggests that Snowden had started to hack into NSA systems and made copies onto pen drives. Of course, his hacking tutorial was essential in this; so did his Sensitive Compartmented Information (SCI) clearance and his access to NSA's Kunia regional base in Hawaii (Epstein, 2017). Without his SCI clearance, he could not have physically accessed the facilities where the NSA's information systems were. The hacking skill ensured him access to the electronic data on those systems, though unauthorized. Thus, Snowden's ability to obtain more documents from the agency was partly because of his hacking skills and the clearance in his possession.

By 2012, Snowden had begun making contacts with famed hacktivists in the The Onion Router (TOR) community. According to Milone (2002), a hacktivist is a person who seeks to bring about social or political changes using technological tools such as TOR and social media. To this end, Snowden contacted Runa Sandvik and Jacob Appelbaum, advocates of TOR, a free and open-source software used to ensure anonymity on the internet by encrypting user traffics through servers (Li et al., 2011; Bartlett, 2015; TOR FAQ, n.d). At this time, we can infer that Snowden's dislike for the NSA and any other corporation that engaged in surveillance was evident. Reaching out to people like Jacob Appelbaum regarding exposing the government's activities undoubtedly spoke about his intentions. Jacob Appelbaum is a well-connected hacker and a representative of WikiLeaks for North America who had made it his aim to use technological tools to elicit social or political change (Milone, 2002; Rich, 2010). Snowden's contact with Appelbaum unequivocally stated his intentions to fight the NSA. To further this cause, he organized a party that provided attendants with knowledge on how to stay anonymous online with software like TOR (Gurnow, 2014; Epstein, 2017). He also set up a two-gigabyte TOR relay server node to contribute to the fight against surveillance (Epstein, 2017).

In all, Snowden's involvement with security agencies from his early days influenced his inclination to work for such organizations. After getting employment opportunities with these agencies, his curiosity led him to learn more about the secret surveillance activities of his employers. His discomfort about these activities did not push him away. Instead, they ignited in him the desire to acquire more information about surveillance programs the security agencies were undertaking. Eventually, his illicit inquisition got him into trouble at the CIA, causing him to unwillingly resign from his position. But fortunately, he landed a new job with DELL, where he continued with his agenda of prying secret information from security agencies like the NSA. His determination led him further to add hacking skills to the IT skills he had, enabling him to take advantage of his position to purloin NSA's secret data from NSANet.

NSA Files and How they Came into Our Hands

While working as a system administrator for the NSA, Snowden began making contacts with Journalists as a way of preparing to go public with his scoop of secret NSA documents (Greenwald, 2014). He contacted some journalists, including Glenn Greenwald, a Guardian newspaper journalist and a litigations lawyer (Greenwald, 2014; Gurnow, 2014; Epstein, 2017). However, due to Greenwald's inability to establish a secure and encrypted means of communication, Snowden halted his communications with him (Gurnow, 2014; Epstein 2017). Earlier, Snowden had asked Greenwald to install encryption software on his computer, but he could not do so (Gurnow, 2014; Epstein, 2017). Perhaps it could have been because Greenwald did not see any good reason to drop his responsibilities to pursue a story from an anonymous informant. Gurnow (2014) later described Snowden's earlier failure to "recruit" Greenwald as a blessing in disguise. Had the communication worked from the onset, Gurnow opines that Greenwald's provocative approach would have taken the spectacle out of the story.

Snowden did not randomly choose Greenwald as the first to help him to break the news of the NSA privacy invasion. His decision was due to Greenwald's credibility in the fight against surveillance-enthusiastic governments (Greenwald, 2014; Kilian, 2014; Epstein, 2017). For instance, Kilian (2014) reveals that Snowden was attracted to Greenwald because, through his articulate writing, he successfully created a community interest that opposed some policies from former President Bush's administration. Besides being a lawyer, Greenwald wrote for an online political magazine called Salon, where he boldly stated his opinions against privacy invasion states (Gurnow, 2014, Epstein, 2017).

According to Gurnow (2014), Greenwald did not only speak against the US government's acts publicly, but he also wrote a book about it. In his book titled *How Would a Patriot Act?* he tirade at the Bush administration for conducting illegal phone tapping and other privacy illegalities. Greenwald leans towards the same libertarian views as Congressman Ron Paul, the politician whose ideologies Snowden aligned with (Epstein, 2017). It is not absurd that the three were connected by the ideology that seeks to fight surveillance-inclined governments. As would be seen later, Snowden and Greenwald pushed forth to make their stance about surveillance clearer.

Despite not getting Greenwald on board through his initial attempts, Snowden did not give up on his plans for him. Instead, he resorted to an alternative to reach his preferred journalist. To achieve this, he went through Laura Poitras, a film activist whom Greenwald, through Salon, had written extensively about how the US government had subjected her to interrogation and searches at airports (Greenwald, 2014). Like Greenwald, Laura spoke against the US government's surveillance programs and exposed their unruly activities in other parts of the world (Glatzer & Poitra, 2006; Greenwald, 2014; Epstein, 2017). It could have been from the article about Poitras on Greenwald's website that Snowden got to know that Greenwald had some connection to Poitras.

On the other hand, Poitras may have been part of Snowden's plan from the onset, considering her bold stance against the US government. But given the unsuccessful attempt to recruit Greenwald, Snowden needed Poitras on board to get him to Greenwald because of his audacity.

By January 23, 2013, Snowden had sent Poitras an email using the alias "Citizen Four" (Greenwald, 2014). Unlike Greenwald, Poitras could get her emails encrypted just like Snowden wanted, giving him the confidence to communicate freely about NSA's documents and programs (Greenwald, 2014; Epstein, 2017). Like Greenwald, Poitras was a member of the Freedom of the Press Foundation (Epstein, 2012; Maass, 2013). She is a graduate of the New School of Public Engagement who went on to take up activist filmmaking as her career (Epstein 2017). In 2006, her documentary movie, "My Country My Country," was released, and in its wake, received several acclamations, including an Oscar nomination. The documentary movie, which brought to the limelight the impact of the Iraq war on its citizens, got her a spot on the US intelligence watchlist (Maass, 2013). Apart from been a target of the US intelligence community for her exposé films, she also engaged in anti-surveillance campaigns with Jacob Appelbaum (Epstein, 2017). She was not new to the realm of working against invasive surveillance.

As has been his focus from the onset, Snowden plunged into the discussions about the dangers of government surveillance with Poitras, citing democracy to be at risk if they did not expose the US government's invasive and unlawful acts (Epstein, 2017). He convinced her further by referring to Stellar Wind, a surveillance program that the NSA during the Bush administration that, without constitutional oversight, authorized the NSA to analyze private and financial transactions of persons in the US (Greenberg, 2014). The Stellar Wind story received a huge outcry from the public when William Binney, a formal NSA Senior crypto mathematician in March 2012, exposed it (Bamford, 2012). According to Epstein (2017), Snowden used the Stellar Wind story to convince Poitras to get on board, but in truth, she was already aware of the invasive activities of the NSA.

Being mindful of the kind of exposé he wanted to bring to the public, Snowden endeavored to get some of the well-known newspapers to help him achieve his goal because of their publishing audacity (Greenwald, 2014). More so, Snowden was convinced about specific newspapers that would back him to achieve the kind of coverage he wanted. He also knew how to get them on board – using the journalists he recruited as his conduits. Greenwald was the guy to help him get the UK-based The Guardian newspaper to publish his work (Epstein, 2014; Greenwald, 2017). Through the efforts of Barton Gellman, her friend at Washington Post's US branch, Poitras would bring the The Guardian newspaper on board (Epstein, 2017). By getting these two giant newspaper agencies on board, Snowden would likely witness the desired results.

According to Greenwald (2014) and Epstein (2017), Poitras contacted her friend who worked with The Washington Post, Barton Gellman, an accomplished journalist who had won several awards, including the Pulitzer Prize award. Gellman also confirms that Poitras reached her in an interview with Michael Kirk (Robertson, 2014). Poitras knew Gellman from the NYU Centre on Law and Security, according to Epstein. As anticipated by Poitras, Gellman expressed interest in the NSA leak. However, he approached the story cautiously. Gellman got the lawyers at The Washington Post to look into the legal implication of handling NSA leaked documents (Epstein, 2017). Undoubtedly, it was a worthy approach on Gellman's path considering the nature of the leaked documents and the implications for those who took part in divulging them.

With Gellman somewhat interested in the story, Poitras moved on to her next assignment, informing Greenwald about the leak and possibly getting him involved (Greenwald, 2014). On April 19, 2013, Poitras arranged a meeting with Greenwald in New York, where he was to deliver a speech for a surveillance program (CAIR-NY, 2013; Greenwald, 2014). According to Greenwald (2014), Poitras told him about the NSA leak Snowden had promised her. All this while, Greenwald did not realize that Snowden was the person who tried to contact him earlier but paused because of the encryption difficulties he faced. He failed to remember because Snowden used different names to disguise himself in the emails, making it difficult for Greenwald to later establish a connection with him (Snowden) on his own. It is also possible that Greenwald may have forgotten about the previous email exchanges. But one thing that may have influenced his approval of the informant was Poitras' involvement, as he trusted her judgment (Greenwald, 2014). The response Poitras hoped for followed as Greenwald agreed to work with them and possibly publish the story in 'The Guardian' newspaper (Greenwald, 2014; Epstein, 2017). After he accepted to work with her, Poitras told him when the document would be ready – mid June 2013. Clearly, Poitras's contact and connection with Greenwald played a critical role in getting him to agree to work on the story.

Regarding the documents Snowden leaked to the journalist, the NSA had a structure for classifying them. The NSA organized the documents into three-level, according to Epstein (2017) and Greenwald (2014). However, other scholars, including King (2013), believe the NSA organized the documents into four (4) groups. The NSA's grouping system classified documents based on the degree of their importance to their activities (King, 2014). More so, the sensitivity of a document's contents determined what category it came under, influencing the security they attributed to it. The NSA devised this approach because they believed some employees would someday become disgruntled and reveal secrets about them (Epstein, 2017). Another reason is to safeguard the documents from employees who may be coerced into revealing the agency's secrets (Epstein, 2017). An agency like the NSA obviously needed mechanisms in place to minimize data breach because of the nature of its work – secret interception of messages, calls, satellite communication, and internet communication (Bamford, 2002). Epstein states that the accessibility of the document was dependent on a need for operation basis, where an employee's duties and positions determine what he or she could access. While it is true that the NSA's system for organizing data helped to mitigate the impact of theft, it also contributed to the overall goal of granting access to employees with minimum risk of divulging secret documents.

Another factor that determined the level a document belonged to was the use the NSA had for it. The levels are 4 in number, and Level 1 contained administrative information such as the Foreign Intelligence Surveillance Act (FISA) court order (King, 2014, Epstein, 2017). In Estein's account, documents at level 2 are typical intelligence that does not contain any details about NSA's sources. He also said Levels 3 and 4 comprised information about how the NSA gained access to their information and contained details of their sources. Epstein also mentions that level 3 documents contained information about stealth projects run by the NSA, CIA, and the Pentagon. According to Epstein, the level 3 documents were referred to as "the Keys to the Kingdom" by one of NSA's executives, and such a description was because of the damaging effect they could have on US intelligence, should the wrong persons get access to them (p. 75). It is difficult to imagine the delicate nature of the last level (level 4), but it probably could contain far more critical information. The levels and their degrees of sensitivities showed how meticulous the agency categorized its data.

The categorization of the documents posed challenges to anyone who wanted to illegally gain access to them. Regardless, Snowden had access to NSA level 1 and 2 data as a Dell employee, but not so for the last two levels (Gurnow, 2014). He had access because contractor companies like Dell had access to only the first two levels; Snowden being an employee of Dell, automatically had access to the first two levels of documents (King, 2014). Snowden also wanted to access the documents in the third and fourth levels (Gurnow, 2014; King, 2014; Epstein, 2017). It is worth mentioning that before Snowden's episode of accessing those high level (levels 3 and 4) documents, he had no authorized access to documents belonging to those levels (Epstein, 2017). But using his position, he accessed the first two data levels through the NSA Network (NSANet) – an official and secret NSA intranet used by CIA, DIA, and NSA employees to share and discuss projects they were working on (Toxen, 2014; Epstein, 2017).

To gain access to the level 3 and 4 documents, Snowden needed to switch to Booz Allen Hamilton (BAH), the company that had access to those documents (King, 2014). BAH is a consultancy firm that handles NSA high-level data (Gurnow, 2014). As a result, Snowden made BAH his new target, using the company as a means to an end. He resigned from his position at Dell as a system administrator in March 2013 for an infrastructure analyst position at BAH (Epstein, 2017). Although his new job position at BAH did not exactly give him access to passwords and the several access privileges he enjoyed at his previous job, he managed to achieve his goal of accessing the level 3 and 4 documents (Epstein, 2017). He was able to break the protocols and the security checks in place to get a hold of the level 3 documents he wanted by March 17, 2013 (Epstein, 2017). Epstein states that not even the several passwords nor the heavily guarded facility could stop him. According to Epstein, Snowden used pre-programmed spider software to capture and index the data he wanted and then copies them onto pen drives.

After obtaining the documents he wanted to hand over to the journalist, he devised a plan to meet them. For in-person meetings with the journalists, Snowden embarked on a trip to Hong Kong (Greenwald, 2014; Gurnow, 2014; Epstein, 2017). Before his travel, he requested a medical vacation from work and left to the said destination on May 18, 2013 (Greenwald, 2014). On his trip, he carried the pen drives that contained the soft copies of the stolen NSA documents (Greenwald, 2014; Gurnow, 2014; Epstein, 2017). Not only did he have the documents ready on pen drives for his Hong Kong trip, but he also had accommodation planned for his time there (Epstein, 2017). Albert Ho, Snowden's lawyer in Hong Kong, also confirmed to Epstein (2017) that someone had arranged accommodation for Snowden before he arrived there.

A few days after he arrived in Hong Kong, Greenwald and Poitras also arrived in the metropolis, and with them was Ewen MacAskill, according to Greenwald. MacAskill accompanied Greenwald to Hong Kong per the instructions of The Guardian’s management to both play a supervisory role and to confirm the genuineness of the story. This approach was a means of assuaging their doubts about the whistleblower. Due to Snowden’s skepticism about who he allowed on the story, Greenwald withheld from introducing MacAskill to Snowden until after he (Greenwald) had met with Snowden to establish the legitimacy of the story (Gurnow, 2014). Soon after, Greenwald introduced MacAskill to Snowden and had him establish Snowden’s authenticity. MacAskill later gave the team, which he was not part of, an assurance that The Guardian would publish the stories fast and with the needed aggressiveness to ensure Snowden’s desired impact (Epstein, 2017). MacAskill’s role was essential, as he pushed for the removal of any bureaucracy that threatened the fast release of the story and helped Greenwald write some of the story (Greenwald, 2014).

In the meantime, Snowden was eager to have Gellman publish his story at his own selected time, but Gellman may have perceived the whole arrangement to be too risky (Gurnow, 2014). Therefore, Gellman approached the story with caution. For instance, he got lawyers at the Washington Post to look into the issues surrounding the handling of NSA leaked documents in Hong Kong after Snowden invited him for a meeting in the China region (Epstein, 2017). Gellman eventually declined to join Snowden in Hong Kong (Greenwald, 2014). Regardless, through Poitras, Snowden provided Gellman with the leaked documents containing an NSA-FBI-CIA collaboration project called PRISM (Greenwald, 2014; Gurnow, 2014). PRISM was a covert mass surveillance program launched under President Bush's administration in 2007 to unjustifiably collect and monitor contents of American residents using data from giant technology companies such as Apple, Yahoo, Microsoft, Google, Facebook, Skype, AOL, and PalTalk (Gellman & Poitras, 2013; Green & MacAskill, 2013). At this point, Snowden needed Gellman on board to get his story out. At least, Snowden wanted Gellman as a backup person, but Gellman's concern for safety got in the way despite the assuring nature of the documents. Regardless, Snowden seems to may have ran out of viable options, causing him to release some of the documents to Gellman despite Gellman's refusal to meet him in Hong Kong.

Snowden did the second distribution when (before his trip to Hong Kong) he directly sent Greenwald a Top-Secret document as a welcome package (Greenwald, 2014). It also was the first time Greenwald received documents from Snowden directly. He later sent Poitras another set of data to be given to Greenwald, however, he instructed Poitras not to hand over the documents to Greenwald until they were in their plane to Hong Kong (Greenwald, 2014; Gurnow, 2014; Epstein, 2017). Included in the documents he sent through Poitras to Greenwald was the Foreign Intelligence Surveillance Act (FISA) warrant that requested Verizon to hand over to the NSA call records of their customers over 90 days (Epstein, 2017). NSA's newsletters were likely in this batch received by Greenwald (Greenwald, 2014). At this point, Snowden had become fully engaged in the business of distributing the NSA documents that were in his possession. But he was cautious not to give all the document to Greenwald, fearing that he might rush to publish it, thereby distorting his plans.

Snowden promised to send another set of the NSA documents to Poitras, but they were encrypted (Greenberg, 2014; Epstein, 2017). According to Snowden's email, the documents would remain encrypted until they had finalized their arrangement (Greenberg, 2014). Possibly, Snowden was referring to them meeting in Hong Kong before he would grant her access to the documents. By this arrangement, Snowden had something to dangle in front of Poitras to ensure she complied to his conditions. Snowden also likely handed over the decryption key to Poitras in Hong Kong, per their agreement. Although that set of documents were encrypted, his manifesto, which was among that batch of documents, was left unencrypted (Greenberg, 2014, Gurnow, 2014).

In retrospect, Snowden may have known that not all the documents he copied were newsworthy. The public did not need to know everything the NSA did. And even if they did, many of those documents, which were highly technical by nature (Greenwald, 2014), would be meaningless to the public, unless those that touch on issues that are easily discernible and directly concern the public. Better still, releasing most of the documents he purloined from the NSA may put into risk the lives of innocent people. So, he only handed over to the journalists documents he thought would interest the public, according to Epstein (2017). In addition, in one of his communications with Greenwald, Snowden said "there is all sort of documents that would have made a big impact that I didn't turn over" (Greenwald, 2014, p. 30). Snowden's assertion that many of the documents in his possession would have made an impact might be true, but given the purported volume of documents he had, it would have been quite difficult for him to go through all those documents before handing them over, as he liked to assess the documents before giving them to the journalists.

Amidst the meetings and distributions of documents to journalists, Snowden knew of the looming troubles that awaited him. Anticipating the consequences that awaited him in the US, Snowden made plans not to return home. He knew without a doubt that authorities in the US would detain him should he have returned to his home country. Also, Snowden claimed the US would have denied him a fair trial if he returned to the US (Fantz, Black, & Martinez, 2013). For this reason, he decided to travel to Ecuador through Russia (Snowden, 2019). In his book "Permanent Record," Snowden (2019) made known that he chose Ecuador because his lawyer indicated that the country and its embassies were favorable to asylum seekers, including Assange. But after landing at the Sheremetyevo airport in Moscow, Russia, things changed. The US State Department stated that they had canceled Snowden's passport (Snowden, 2019). The passport situation and the pressure from the US made it impossible for him to leave the Russian airport (Snowden, 2019). After fourth days at Sheremetyevo and many failed attempts to leave Russia, the government eventually granted him a one-year temporal asylum, allowing him to leave the airport on 1st August 2013 (Loiko, 2015; Snowden, 2019). It is worth mentioning that Russia had no extradition treaty with the US and was not bound to lease him (Hildebrandt, 2013). After a series of his residence renewals in Russia, however, in October 2020, the then 37-year-old whistleblower received his permanent residence in Russia (Ilyushina, October 2020). In a tweet in November 2020, Snowden announced that he was applying for dual US-Russian citizenship. Currently, Snowden and his wife are in Russia with their newly born (Gordon, 2020). Snowden's attempt to leave Hong Kong to Ecuador failed, leaving him in Russia. But ultimately, he was able to secure asylum in Russia and citizenship in the country where he currently is with his wife and their son.

Regardless of Snowden's personal life, some of the journalists made it their goal to share soft copies of the documents accessible by the public. To achieve this goal, Greenwald, for instance, worked with The Intercept to publish on their website some of the leaked NSA documents they taught would be of interest to the public. The Intercept, a Pierre Omidyar funded website, is a news organization that endeavors to fight corruption, criminal justice, war, surveillance, among other malicious acts through bold analysis of stories, legal, and editorial support for their journalist. While other journalists and the public may have circulated these leaked documents over the internet, including on GitHub, I obtained this scoop of newsletter documents (mostly in PDFs) from The Intercept (About the Intercept, 2019). I downloaded these PDFs from them because I believe they have the most comprehensive set of documents.

Although over two thousand copies of the NSA's newsletters exist in the public domain, there are speculations about the number of documents Snowden stole from the NSA. In an interview with King, McConnell claimed that the number of documents Snowden purloined was between 1.7 to 1.8 million (King, 2014). In Epstein's account, Snowden stole over one million documents from the NSA. Other people have estimated figures that are significantly lower than that by McConnel. In one of his posts about Snowden exposé, Alexander Keith, former NSA director, revealed that Snowden took between 50,000 and 200,000 documents (Zaleski, 2013). In an interview with James Bamford, Snowden himself denied that the number of documents in his possession was anything close to the 1.7 million number circulating in the media. He, however, refused to disclose the number (Watson, 2014). Regardless of the number of documents he got his hands on, one thing was for sure -- he managed to get a hold of enough data to tarnish the image of the NSA in the public space. Although most of the materials that made waves in the media were dominantly level 1 (administrative level) documents, he managed to make the most of them. Some people may speculate that the remaining documents could reveal more about the NSA to the public, but that would not be relevant to the masses. Most of those documents would likely be technical and not as easily comprehensible as those in level 1, which by essence also required some interpretation, although minimal.

The tasks involved in sharing NSA documents with the journalists were not easy. Apart from the legal implications of their actions, Snowden and his colleagues faced odds that could have jeopardized their chances of making the news available to us; they persevered, working tirelessly to overcome journalistic bureaucracies and the surveillance they strived to expose. But finally, they prevailed in not only bringing to the public nerve-chilling stories about the NSA's activities, but they also managed to bring us some of the documents in PDFs and on The Intercept's website where anybody in the world could access them. Among others, Glenn Greenwald, Laura Poitras, and Ewan MacAskill were particularly instrumental in exposing the NSA as they worked with Snowden to release the stories and documents in impactful ways. They also provided Snowden with some legal advice on the implications of these actions. Because of their combined efforts and collaboration, we have in our possession NSA's newsletter documents to carry out our study about the agency's culture.

Study Rationale

Most of the knowledge we possess about the NSA comes from speculators or researchers with a keen interest in the agency. While the knowledge these speculators and researchers provide may, in many respects, be accurate, there exist the possibility of them misrepresenting some facts. Fortunately, we have access to NSA's newsletters. Primary sources of information such as the NSA's newsletters typically provide better accuracy on issues. Therefore, we now have the opportunity to assess or evaluate the information speculators and the history books have offered.

In assessing these newsletters, there are things we need to bear in mind. Thus, while analyzing these newsletters may reveal details similar to what we already know, including the ones on Edward Snowden and the NSA's history, the fact remains that this thesis would present a perspective that would either corroborate or refute knowledge we already possess about the agency. Likewise, this analysis may also bring forth the knowledge that previous works could not have unraveled, adding to the overall literature we possess about the NSA.

Chapter Two: Methodology

To understand better the NSA's topics (themes), I used Content Analysis and Natural Language Processing (NLP) techniques in which topic modeling plays an important role. Also, I utilized Key Word in Context (KWIC), an indexing approach, and a concordance format that allows humans to efficiently search through text in a way that enables it to be surrounded by the text it originally appeared with to a certain degree (Miller et al., 1993). To start, I have segmented this chapter into two – Data Preparation and Data Analysis.

Data Preparation

I outline below the steps needed to prepare the data used in this study.

a. Description of the data (SIDtoday)

b. Data Gathering

c. Extraction and preprocessing of data

a. Description of the data (SIDtoday)

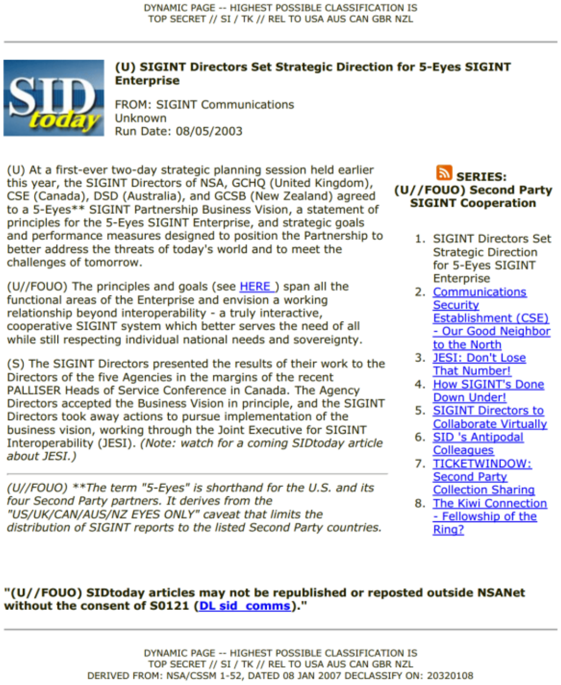
In his article titled *What It's Like to Read the NSA's Newspaper for Spies*, Peter Maass (2016) described SIDtoday as an internal online newsletter publication created by the communication team of the Signals Intelligence Directorate (SID) within the NSA to keep its employees up to date on issues about the agency and intelligence activities in general. To achieve the goals of informing their employees, the agency introduced SIDtoday, a news website that operated within a highly secured and classified network was launched on March 31, 2003 by the NSA (Maas, 2016). In this secured SIDtoday environment, employees could read the news, interact among themselves about projects, and share community-specific information. According to the SIGINT communication team of the NSA, the goal of the new internal online website (SIDtoday) was to centralize communications of the SIGINT directorate, making it easy to keep employees updated on issues within the Signals Intelligence Directorate (SID) (Welcome to SIDtoday, 2003). The introduction of this website and its emphasis on a central communication channel imply that SIDtoday replaced a less effective medium that probably lacked the improvement this new one has. Those responsible for SIDtoday may have perceived the absence of a central system for information dissemination within the community as a hindrance to a smooth flow of information. Therefore, a centralized news website may have been one of the most viable approaches for addressing the challenges of disseminating information within the agency.

In terms of numbers, SIDtoday's website contained a sizeable amount of news articles for its readers. Since its launch, readers of its contents had access to a pool of several articles, ranging from lackluster to astonishing contents that revealed both the spectacular and the mundane nature of the agency and its employees (Number 9, 2012). From this collection of articles, The Intercept made available to the public newsletters that Edward Snowden exposed. From his exposé, we could guess that annually the agency published an average of 500 news articles on SIDtoday. Also, it indicated that on many occasions, the agency published at least one article on the website daily, indicating that there was almost always something on the website to read. Regardless, there were days when the agency published no news item on SIDtoday.

Regarding appearance, although we do not have access to SIDtoday's website, we can tell from the PDFs that it had a typical website look from the arrangement of the images and the single-spaced text on the pages. Figure 1 provides a snapshot of the PDF. One would also notice that most of the articles are half-page filled.

Again, regarding the SIDtoday's appearance, we can tell that the webpages are in three segments: header, body, and footer. First, in the header segment, you see the header text "DYNAMIC PAGE – HIGHEST POSSIBLE CLASSIFICATION IS TOP SECRET // SI / TK // REL TO USA AUS CAN GBR NZL." The text DYNAMIC PAGE suggests that the NSA frequently publishes content on the websites. USA, AUS, CAN, GBR, NZL are abbreviations for the United States of America, Australia, Canada, Great Britain, and New Zealand, resectively. The remaining text, such as TOP SECRET, suggests the sensitivity of the newsletter. Note that the header text runs throughout all the publications. The body, the second section, has at its top-left area, the logo of the newsletter. To the right of the body segment, you will see the newsletter's title, the author's name, and the publication name. You may also observe that article titles are prefixed by letters in parenthesis. These letters indicate the security designation of the newsletter. For instance, in the newsletter for April 1, 2003, titled Practical Jokes and April Fool, is prefixed by (U), indicating it is unclassified.

Figure 1.
Screenshot of SIDtoday's newsletter in PDF



DYNAMIC PAGE -- HIGHEST POSSIBLE CLASSIFICATION IS
TOP SECRET // SI / TK // REL TO USA AUS CAN GBR NZL

(U) SIGINT Directors Set Strategic Direction for 5-Eyes SIGINT Enterprise

FROM: SIGINT Communications
Unknown
Run Date: 08/05/2003

(U) At a first-ever two-day strategic planning session held earlier this year, the SIGINT Directors of NSA, GCHQ (United Kingdom), CSE (Canada), DSD (Australia), and GCSB (New Zealand) agreed to a 5-Eyes** SIGINT Partnership Business Vision, a statement of principles for the 5-Eyes SIGINT Enterprise, and strategic goals and performance measures designed to position the Partnership to better address the threats of today's world and to meet the challenges of tomorrow.

(U//FOUO) The principles and goals (see HERE) span all the functional areas of the Enterprise and envision a working relationship beyond interoperability - a truly interactive, cooperative SIGINT system which better serves the need of all while still respecting individual national needs and sovereignty.

(S) The SIGINT Directors presented the results of their work to the Directors of the five Agencies in the margins of the recent PALLISER Heads of Service Conference in Canada. The Agency Directors accepted the Business Vision in principle, and the SIGINT Directors took away actions to pursue implementation of the business vision, working through the Joint Executive for SIGINT Interoperability (JESI). *(Note: watch for a coming SIDtoday article about JESI.)*

*(U//FOUO) **The term "5-Eyes" is shorthand for the U.S. and its four Second Party partners. It derives from the "US/UK/CAN/AUS/NZ EYES ONLY" caveat that limits the distribution of SIGINT reports to the listed Second Party countries.*

**SERIES:
(U//FOUO) Second Party SIGINT Cooperation**

1. SIGINT Directors Set Strategic Direction for 5-Eyes SIGINT Enterprise
2. Communications Security Establishment (CSE) - Our Good Neighbor to the North
3. JESI: Don't Lose That Number!
4. How SIGINT's Done Down Under!
5. SIGINT Directors to Collaborate Virtually
6. SID 's Antipodal Colleagues
7. TICKETWINDOW: Second Party Collection Sharing
8. The Kiwi Connection - Fellowship of the Ring?

"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 (DL sid comms)."

DYNAMIC PAGE -- HIGHEST POSSIBLE CLASSIFICATION IS
TOP SECRET // SI / TK // REL TO USA AUS CAN GBR NZL
DERIVED FROM: NSA/CSSM 1-52, DATED 08 JAN 2007 DECLASSIFY ON: 20320108

This figure gives us a view of SIDtoday in its original format.

Furthermore, in the body section, you may observe a couple of paragraphs, bullet points, or images as contents. At the very bottom, you will find, running across all the published articles, a boldfaced text that reads *SIDtoday articles may not be republished or reposted outside of NSAnet without the consent of S0121*. Below this text, you will see the footer, the third section. The third section, which also is the last section, has some similarities to the first section (header). However, this third section contains additional information, which includes the article's classification date. The text structure is such that the header, body, and footer sections give SIDtoday a standard interface that woud be familiar to most readers of newspapers or website contents.

When it comes to content, SIDtoday contains a wide variety of discussions. For instance, it contains articles about projects that NSA's employees have been on, their approach for working on them, and the reasons for undertaking them, according to Maas (2016). As well, it had stories of important events about the agency that were written for amusement and, in some cases, to teach lessons or both (Practical Jokes and April Fools., 2003). Stories about the previous jobs of the agency's employees were also in SIDtoday, as well as the work experiences of the agency's interns. Further to the topics SIDtoday covered, some publications provided details of what officials at the agency were reading, including novels and technical documents they were reading at the time (Maas, 2016). SIDtoday publications about what some leaders in the Signal Intelligence Directorate were reading at the time, to a large extent, give us an insight into that culture of openness in this community. But in a general sense, SIDtoday suggests to us the wide range of topics it provides its readers.

In terms of characteristics, the contents of SIDtoday reveal certain features that are either common or peculiar to the agency, giving us a glimpse of the organizational culture within the agency. Thus, some cultural elements of the agency can be perceived by reading their newsletters. For instance, some of the vocabularies that inform us of the NSA's culture are evident in the agency's April 22, 2003 newsletter called *Dynamic Methods of Interaction with New and Existing Customers*. Here, SIDtoday reveals that the agency refers to departments or agencies that rely on it for intelligence as customers while referring to intelligence as product (Dynamic Methods of Interaction with New and Existing Customers, 2003). Similarly, the publications divulged that the agency refers to objects and projects using terminology such as Anchory, a database for keeping records of intelligence reports and analyses (Nash, 2013). Although not comprehensive, the vocabularies, among other things, helped to inform us of some of the cultural elements within the NSA.

Next to its cultural pointers is its authorship. From the contents of some of the articles, it appears the editors of SIDtoday welcome the agency's employees to make submissions for publication on the website. Also, some employees have taken up writing in SIDtoday as their side gig. There also are instances where SIDtoday features guest columnists to write content for the website (National-Cryptologic-School, 2006). Also, judging from the contents on SIDtoday's pages, we can tell that persons using pseudonyms such as SIGINT Curmudgeon, Zelda, and Abby are among those who write in the newsletter.

Additionally, there are other groups of people who get the opportunity to publish in SIDtoday. For instance, person who did not work in the communication team got the chance to write about his employment life before joining the NSA ("Odd Jobs Before NSA," 2005). Likewise, others, including interns at the agency, receive permission to publish in the newsletter, detailing their experience at the organization (Culture Shock. 2011). Another thing that widens the authorship is the "Questions and Answers" sections featured in the newsletter. Here, high-ranking officials of the agency wrote about their responsibilities and the motivation for their actions. Indeed, the different variety of people who publish in SIDtoday is evidence of the communication unit's diverse authorship approach.

Concerning frequency, SIDtoday churns out articles frequently. Mostly for the articles published from 2003 to 2006, the daily publication numbers ranged between 1 and 3. However, some days had no publications. We note that *The Intercept* made available more articles for these years than the remaining 2007 to 2012. These numbers give us a good indication that SIDtoday received a regular publication.

In terms of uniqueness, SIDtoday has some distinctiveness, including its confidential nature. According to Maas (2016), this confidentiality is because of the sensitive nature of the newsletters. Due to SIDtoday's sensitivity, any article that provided details of projects and how they were/are conducted, without any doubt, requires protection from unauthorized persons. To ensure confidentiality, only workers of the NSA could access the newsletter (Maas, 2016). The secrets of the NSA, among other things, give the agency a unique approach to issues.

In summary, on March 31, 2003, the online portal called SIDtoday was introduced by the NSA as a hub to make information easily accessible to NSA's employees. As a web resource, SIDtoday's goal was to serve as a centralized news website. In addition to its "parlance" interface, the various topics covered on its pages reflects the agency's endeavor to touch on several facets of issues that may interest its exclusive readers. As well, the "Question and Answers" sections are evidence of SIDtoday's interactive nature, proving they welcome employees and interns to publish on SIDtoday through a number of ways. Security-wise, the newsletter is accessible only internally and by employees of the NSA. The trove of information in SIDtoday makes it a valuable source of knowledge about the agency.

b. Data Gathering

The data (newsletter articles) used in this research were downloaded from https://theintercept.com/snowden-sidtoday. SIDtoday is typically published for a high-level department or units in the NSA (Snowden Archive, n.d.). Therefore, access to these newsletters gives this study the privilege to gain insight into some useful discussions within the said US intelligence community.

Available on The Intercept's website – www.theintercept.com – are over 2,000 editorially reviewed newsletter documents of the NSA's, spanning from 2003 to 2012 (Snowden Archive, n.d.). The newsletters were present both in a portable document format (PDF) and in a text format, with each batch sorted according to their dates of publication.

To gain access, you could view the newsletters online or download them. And to download them, one option was to get the periodically released batches by *The Intercept* through their website. The other download option was to obtain them from GitHub, a social platform for coding and repositories that offer a distributed means of controlling versions of code using a feature called gits (Russell, 2013). Unlike the download options on *The Intercept's* website, the newsletters available for download on GitHub were properly organized, as all the documents were grouped in different folders and sorted according to their years of publication. In addition to presenting the newsletters in compressed or archives files, they labeled all the newsletters in folders using a consistent naming convention (date-title) that makes searching by dates easy.

To ensure I obtained enough relevant data for this study, I collected all the newsletters available on the GitHub repository – https://github.com/firstlookmedia/sidtoday/ --. As seen in Table 1, I used for this study the newsletters from 2003 to 2006 and ignored those from 2007 to 2012. My decision to use newsletters from 2003 to 2006 was because of the insufficient newsletters available for the other years. For the years 2007, 2008, 2009, 2011, 2012, the number of newsletters available was 7, 4, 7, 2, and 1 respectively, indicating that all those years could only contribute 21 newsletters to the study, which would have had an insignificant impact on this topic modeling study. Despite the scarcity and difficulty associated with obtaining documents of this kind, removing the years with inadequate text helped to avoid errors that would possibly arise from a small sample size. Limiting the study to periods from 2003 to 2006 was ideal because those years contained enough articles (an average of 530 per year) for this analysis, whereas those from 2007 to 2012 provided inadequate newsletters. Although the number of SIDtoday articles available to the public is scanty compared to the speculated amount in Snowden's possession, adding those from years with insufficient newsletters would not prove beneficial to this study.

Table 1. SIDtoday newsletters according to publication years.

Year	2003	2004	2005	2006	**Total**
Number of Newsletters	415	517	596	593	2121

c. Extraction and Preprocessing

Data preprocessing is a prerequisite in data mining, and it aims at preparing text for data analysis (Bhatia, 2019). To conduct this prerequisite, I followed the data gathering process with preprocesses techniques to ensure the readiness of the data for the analysis.

First, I decompress or unzip the archived newsletters from its originally compressed state. Thus, the data I downloaded from The Intercept came in a compressed format (zip files) and needed to be uncompressed. Without the decompression process, I could not have processed the data further.

To decompress the data, I used a macOS software called The Unarchiver. I ran the 74-megabyte zipped file through The Unarchiver, which produced an 85-megabyte file as its output and contained the various newsletters grouped according to their years of publication. There were nine (9) folders in the decompressed folder, and they span from 2003 to 2012. I, however, did not find newsletters for 2010 in the decompressed file.

Next, I obtain the text data from the PDF newsletters using the Python programming software and the publicly available PDFMiner library. For this step, I follow the methodology for extracting text from PDF in Aggarwal's (2005) and Shinyama's (2019). I extracted the raw text from the PDFs. However, I was unable to extract from a couple of the PDFs the raw text, as they were in formats that did yield to my text extraction approach.

To obtain further the text data from the three (3) PDF newsletters that my previous codes could not process, I utilized a different python library and set of new python scripts. In the new code scripts, as part of other python libraries, I used the Pytesseract library, an optical character recognition (OCR) python library capable of reading text in images (Lee, 2020). I used this library because texts in images can effectively be recognized and extracted by OCR programs like Pytesseract. Using this new library and the new scripts, I was ultimately able to obtain the raw text from the three remaining PDFs.

Following the OCR process, I compiled the 2121 newsletters for further data preprocessing. Comprised in this data preprocessing phase are data cleaning and feature extraction from the data (Aggarwal, 2015). I needed to clean the 2121 text articles, removing all irrelevant characters to the research (Stein & Eissen, 2004; Guo et al., 2016). For instance, the header and footer sections of SIDtoday, which contain information on security designations for the newsletters, were of no relevance to most of the analysis that I undertook in this research. More so, the text in the header and footer sections are similar for all the newsletters. The irrelevant texts that I removed include metadata such as title, dates, author/s, footers, header, and newsletter designation labels. Freeing the text data from these irrelevant characters is necessary because it mitigates anomalies in the text data and subsequently in the outputs or results. Also, closely previewing the newsletter helped me to ensure that I was targeting the main body of text data for my study with minimal intrusion from noise.

Regarding the extraction, I selected at random some of the newsletters to manually scan (using close reading) and familiarize myself with the text, interact with them, and identify those sections that were relevant to my study. Thus, I needed to obtain certain elements or attributes from the newsletters that I consider relevant for the research work. According to Aggarwal (2015), this extraction process involved is described as feature extraction. Feature extraction is a process that concerns itself with identifying and retrieving parts of the raw data that are relevant to a study (Aggarwal, 2015: Sarkar, 2019). I consider the feature extraction step useful, as this study only needed certain portions of SIDtoday's articles, ignoring those that were irrelevant to it (Guo et al., 2016).

Besides, I familiarized myself further with the newsletter by identifying other relevant components. This identification was in the form of ascertaining metadata contained in the newsletter such as: name of the writer or contributor(s) label as FROM; the position of the contributor(s); the name of the contributing unit; run date; and classification and dissemination designations of the newsletter expressed either as Unclassified(U), Confidential (C), Secret (S), Top Secret (TS), No Foreign Nationals Distribution (NOFORN), For Official Use Only (FOUO), Sensitive Compartmented Information (SCI), or Release To (REL TO) (Intelligence Community Classification Guidance Findings and Recommendations Report, 2008). By ascertaining this information, I got a better understanding of the newsletters' structure, as well as familiarize myself with those that are important to my research.

Interacting or reading through SIDtoday played another important purpose, as it made me aware of the need to find and utilize a tool that could further clean and select the data I needed for this study. The tool I needed for this task had to be able to match patterns in text. As such, the tools I identified include Regular Expressions (regex). Regex is a program that can match patterns in text (Nield, 2019). I used it not only to clean the newsletters but also to select and extract parts of their content that are important to this study. Without interacting, I would not have identified some of the tools I required, including regular expression.

Consistent with the work of Rockwell and Sinclair (2016), and in their book -- *Hermeneutica*, I had to extract or separate parts of the text for the various analysis I needed to conduct. For instance, I had to separate the body content of all the SIDtoday's articles from the full text, omitting metadata like the title date, author, and any other text I considered irrelevant. To do this isolation and extraction, again, I found regex to be the most appropriate tool. That said, regex played an important role in allowing me to capture the body section of the newsletters.

Following the pattern matching stage, I needed to organize the text from the newsletters to enable efficient processing. Therefore, I decided to segment and group the various text. For the segmentation and grouping of the text, as Rockwell and Sinclair (2016) did in their analysis of David Hume's Dialogue, I used short yet descriptive names to keep track of the various versions of texts I created from the originals. For instance, I grouped into four (4) folders all the newsletters according to their respective years. I arranged the raw text files within these four (4) folders, further sorting them according to their months of productions. This way, I had folders ranging from 2003 to 2006. On average, each folder contained about 530 newsletters. From here, I began to refer to the text as my data.

Ultimately, I had 2121 articles and categorized them appropriately for the various studies I undertook. For Data A, I segmented the data into their respective years of publication, giving me four (4) folders, each containing 530 articles on average. As well, I combined all the text data into one file and labeled it Data B. Thus, I sectioned the corpus into two different categories and labeled them Data A and B.

In terms of errors, SIDtoday's newsletters contained minimal grammatical inaccuracies. Unlike text obtainable from social media platforms, like Twitter, which is typically prone to grammatical errors and spelling mistakes (Guo et al., 2016), the newsletters proved otherwise. The mistakes are less in SIDtoday's publications because of checks the agency's editorial unit, the SIGINT Communication team, had implemented to ensure their articles undergo a series of editing before publication. Evidence of the presence of an editorial unit in the NSA is in their November 19, 2003 article *Talk the Pen by the Horns*. In the article, the writer invited readers to become guest editors of SIDtoday, suggesting the presence of an editorial unit. That said, the presence of an editorial body that checks SIDtoday's contents assures me that the data I used in this research were written correctly. This fact frees me from the need to conduct spell and grammatical checks. Randomly reading the newsletters also helped me to corroborate that the newsletters indeed contained minimal errors. Due to the well-constructed nature of the data, I will be working on within this study, I omitted the need to undertake a grammatical and spelling check correction. Most importantly, one of the algorithms (Latent Dirichlet Allocation) I used for this study is, according to the works of Tang, Zhang, and Mei, (2013), more effective in detecting topics in a text that are well-written. Thus, the grammatical accuracy of SIDtoday’s content did not only free this study from the extra step but also contributed to obtaining more accurate results from the LDA model that I will use for the analysis.

In all, the preliminary steps involved in the preprocessing stage have proven useful, resulting in uniform sets of text data – Data A and B. Data A comprises four (4) folders, and each contains the newsletters for that year. Data B contains one folder containing all the articles combined -- 2121 articles. Before categorizing the data, the processes that I engaged in proved useful too. These processes include unzipping the newsletters and converting the PDF to text. As well, the removal of some metadata and other non-essential characters were much needed preprocessing steps. This cleaning process provided me with the basis for proceeding with the analysis.

Analysis

Knowing that the NSA is involved in gathering Signal Intelligence is, for me, not enough. Instead, this knowledge motivates me to learn as much as possible about the agency. Therefore, I wanted to, from their SIDtoday, obtain more knowledge about their activities regarding how they handle affairs internally. As such, I had to probe SIDtoday's newsletters to provide answers to the following questions.

a. What are the common topics or themes that come up in the NSA's newsletters?

b. Do any of the topics in NSA's newsletters (SIDtoday) relate to data surveillance? What are their labels?

c. What trends do we notice about NSA's data surveillance topics?

a. What are the common themes and issues that come up in the NSA's newsletter (SIDtoday)?

To ascertain the common issues that the NSA discusses in SIDtoday, I used a method called topic modeling, as it is an efficient means of deriving themes from unstructured text data.

Topic modeling

Topic modeling (TM) is a text analysis algorithm that uses statistical methods to examine words to uncover topics or thematical structures in a collection of documents, also called corpus (Blei, 2012; EMC Education Services, 2015; Leydesdorff, & Nerghes, 2017). On the other hand, topics refer to a set of vocabulary with distribution over the words (EMC Education Services, 2015). Blei (2012) defines topics as themes in a collection of documents. That said, the topic modeling algorithm creates sets of themes from text data through its analysis. It operates by revealing hidden topical patterns in a corpus, annotates the text, and then organizes the words (tokens) according to the annotations. Rish et al. (2014) have described topic modeling as a data-mining tool that is capable of capturing topics from an unordered group of features, usually words. Although supervised in some cases, TM typically functions as an unsupervised algorithm, making it independent and free from external influences (Leydesdorff & Nerghes, 2017). TM usually does not require any prior knowledge of the document as it can generate the topics from analyzing the corpus.

Topic modeling's usefulness, for long, has been prevalent in the computer science field until recently when Blei (2012), a computer scientist and developer, introduced to humanists and social scientists the usefulness of this computer algorithm in discovering topics or themes from large corpora. Using his TM algorithm, Blei (2012) unveiled about 100 topics from 17,000 scholarly scientific articles, demonstrating to scholars of other fields, particularly humanists and social scientists, how they can use this tool in their studies. TM's use extends to allow us to explore, summarize, visualize, organize, search, and theorize about a corpus (Blei, 2012; EMC Education Services, 2015). Currently, in addition to the humanist and social scientists who utilize TM for their research works, some businesses have started using it to analyze text data from various sources, including emails, social media platforms, and the Internet of Things.

In terms of the way it works, TM uses some models and mechanisms for its operation. Using computer models such as Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Probabilistic Latent Semantic Analysis (pLDA), TM assumes that documents about similar concepts will likely contain similar terms (words). Blei simplifies this assumption by arguing that "terms that frequently occur together tend to be about the same topic." (Blei, 2012, para. 9). In simple terms, TM counts words and groups those with a similar pattern to infer topics in text data. For instance, a document about fish and meat will contain words that you are likely to see associated with these topics. Therefore, topic models operate by revealing these words in clusters that represent them in their respective proportions. For instance, in a document about fish, you are likely to see terms like these: aquatic, mackerel, water, river, fillet, fin, net, nibble, fingerling, catch, etc. Similarly, a document about meat may contain words like beef, chicken, meaty, butchery, pork, barbecue, steak, etc. These terms or themes are brought forth from the corpus by the topic modeling algorithm to inform on what a document is talking about.

Using computer algorithms, topic modeling can analyze various sets of text to reveal thematical and topical patterns, aiding researchers in understanding a corpus without having to engage in a word by word reading. Among others, the Latent Dirichlet Allocation (LDA) algorithm that topic modeling uses, in addition to the theme clustering assumptions that underline its operations, makes topic modeling a useful tool for people who work with large volumes of text. Through initiatives by people like David Blei, topic modeling is now part of the toolkits of humanities and social scholars.

Adapting the Data for topic modeling

Using the Python programming language, I wrote a set of scripts to clean further Data B in a manner that is well suited for topic modeling. Data B is the corpus that contains the 2121 newsletters in a single folder. This cleaning process is an addition to that which I conducted at the preprocessing stage. Also, I tailored it to expunge text that I consider irrelevant to the topic modeling.

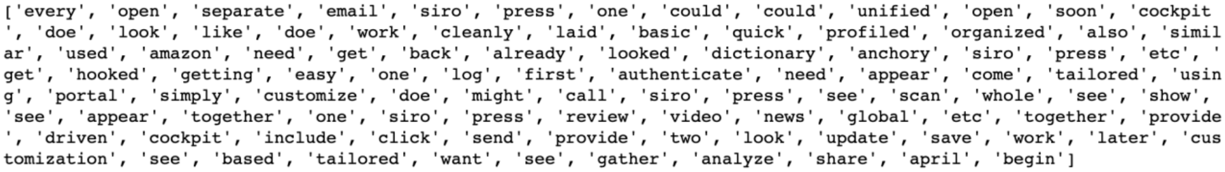
The first step involved in the cleaning process after I imported the data was to lemmatize the text. I followed that process by changing all the letters to their lowercase. According to Toman et al. (2006), lemmatization is the process of removing inflections from words and retaining only their dictionary forms or base. The removal helped me to reduce the number of words (features or dimensions) in the newsletters, ultimately improving the results the analysis produces. In addition to lemmatizing the corpus, I removed all irrelevant elements from the data (corpus), including all stopwords. I did the removal by combining the stopwords libraries of both the NLTK and Spacy, as the combination provided an extensive list of words to use.

Also, I customized the combined stopwords by adding terms present in the corpus that did not have any semantic value to the topic modeling. These words or characters included SI, TK, REL, USA, AUS, GBR, NZL, U, SIRO, SSI, SIGINT, NSA, TSI, SIG, (S//SI), (TS//SI), (U), and (U//FOUO). For some of these characters, SID included them to inform readers about the security level of the newsletter or the designation of a particular article. For others like SID, SIGNIT, and NSA, I added them to the stopword list because they would typically have a high frequency in the corpus but would not be semantically relevant to our topic models. Therefore, removing the stopwords from the data was helpful, as many could have impacted the results negatively. Ultimately, I expunged the data (corpus) of any text or word that is a stopword or NSA's newsletter designation text.

After the customized removal of some texts, I cleaned the data of all punctuations and excess whitespaces. Since punctuations and whitespace are necessary for every text that is supposed to be comprehensible to human, it reasonable to expect that NSA's newsletter, because its goal is to communicate to humans and do so clearly, would have many punctuation marks (Clark, 2003; Minnich et al., 2016). Despite the vital role of punctuation marks in the newsletter's comprehensibility, they, unfortunately, do not provide any semantic meaning on their own. As such, I removed all punctuations and all non-single whitespaces before I ran the corpus through the model, as it made processing the corpus by the algorithm easier.

Next, I did a part of speech (POS) tagging to narrow down to the most relevant text in the corpus. POS is a grammatical process that marks words with their corresponding parts of speech such as noun, verb, adjective, adverb, pronoun, preposition, conjunction, interjection, numeral, article, and determiner (Ratnaparkhi, 1996; Gimpel et al., 2010). Following Schofield et al. (2016) recommendation, I tagged the corpus according to their various speech categories and removed text that provided little thematic meaning to the model while leaving those that are semantically relevant such as nouns, verbs, and adjectives. The POS tagging process also helped me remove noise the previous processes could not remove. Thus, by limiting the corpus to adjectives, nouns, and verbs only. See figure 2 for a sample of the preprocessed output.

Figure 2.
Sample of the preprocessed text data



```
['every', 'open', 'separate', 'email', 'siro', 'press', 'one', 'could', 'could', 'unified', 'open', 'soon', 'cockpit
', 'doe', 'look', 'like', 'doe', 'work', 'cleanly', 'laid', 'basic', 'quick', 'profiled', 'organized', 'also', 'simil
ar', 'used', 'amazon', 'need', 'get', 'back', 'already', 'looked', 'dictionary', 'anchory', 'siro', 'press', 'etc', '
get', 'hooked', 'getting', 'easy', 'one', 'log', 'first', 'authenticate', 'need', 'appear', 'come', 'tailored', 'usin
g', 'portal', 'simply', 'customize', 'doe', 'might', 'call', 'siro', 'press', 'see', 'scan', 'whole', 'see', 'show',
'see', 'appear', 'together', 'one', 'siro', 'press', 'review', 'video', 'news', 'global', 'etc', 'together', 'provide
', 'driven', 'cockpit', 'include', 'click', 'send', 'provide', 'two', 'look', 'update', 'save', 'work', 'later', 'cus
tomization', 'see', 'based', 'tailored', 'want', 'see', 'gather', 'analyze', 'share', 'april', 'begin']
```

When I completed the above processes, I reviewed a couple of topic modeling algorithms and their implementation methods. The algorithms I reviewed include Latent Semantic Indexing (LSI), Probability Latent Semantic Indexing (PLSI), Latent Dirichlet Allocation (LDA), and Non-Negative Matrix Factorization (NMF) (Blei, 2012). All these algorithms proved useful to my purpose of identifying themes in my corpus. However, I selected LDA because of its efficiency, availability of documentation, and ability to generalize to other documents (Xu, 2019). As well, LDA is also able to infer more accurately for unknown documents than its counterparts.

Tools

Like Evans (2014), before making my choice on which tool to use, I reviewed several topic modeling implementation software and packages. The software I reviewed included the R programming language, Python programming language, and MALLET (Machine Language for Language Toolkit). While all these were excellent tools for my topic modeling needs, I opted for the Python programming language as my preferred tool because, in addition to its ability to produce reproducible results, it offered many visualization features for interpreting the results. More so, its processing time was relatively lower as opposed to those of its competitors. Also, it is highly customizable, as it not only allows you to alter few parameters but enables you to modify several, including the hyperparameters in topic modeling. As well, Python offers access to several algorithmic packages – Gensim, Scikit-Learn, and NLTK – for implementing topic modeling for text data (Sarkar, 2019). More so, it makes provision for packages like MALLET.

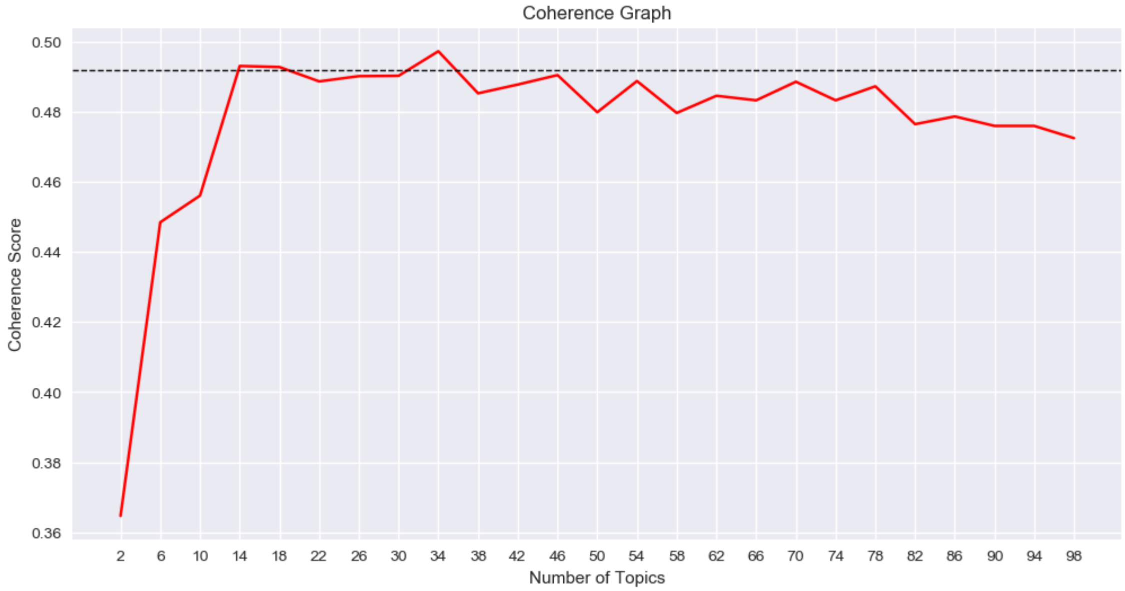
In selecting the ideal package implementation for my study, I opted for MALLET. Developed by Andrew McCallum and his colleagues, MALLET is a Java-based package that is implemented in the Python programming language through a Gensim wrapper (McCallum, 2002; Barde & Bainwad, 2017). This natural language processing (NLP) toolkit serves as a useful tool for text classification, text clustering, information retrieval, topic modeling, and other text mining tasks (McCallum, 2002; Liu et al., 2016). In addition to it being open-source, MALLET supports implementations such as the Latent Dirichlet Allocation (LDA), Hierarchical Latent Dirichlet Allocation (HLDA), and Pachinko Allocation Model (PAM) (Liu et al., 2016). Unlike the other topic modeling packages, MALLET is easy to use and offers a graphical interface for non-coding researchers. Through my evaluation, I noticed that MALLET's benefits and the Python programming language leverages them over their alternatives, influencing my decision to make them my preferred tools.

Configuration of MALLET

Here, I first downloaded the MALLET package on to my computer and referred to it in Python. From there, I went on to set the parameters for the MALLET. Unlike parameters such as "Iterations" and "workers," the one I spent the most time on was the "Number of Topics," as it is the most important in any topic modeling algorithm (Greene, Callaghan, & Cunningham, 2014; Zhao et al., 2015). In determining the optimal number of topics to use for the model, I followed the statistical approach recommended by Mimno et al. (2011) and Sarkar (2019). These scholars have indicated that the number of topics for an LDA model can be optimized by utilizing the coherence or the perplexity test score. It is worth mentioning that my choice of LDA was because its results were more intelligible than those of its contenders. Therefore, I used the coherence score test approach to ascertain the score each LDA model produced. In computing the coherence scores, I ran the model on topic numbers between 2 and 96 and in multiples of 4. Thus, I tested 25 topics, ranging from 2 to 96, with each being a multiple of 4 to obtain the coherence score.

After computing the coherence test with the various topic numbers, I obtained coherence scores that ranged between 0.31 and 0.47 for the entire corpus, giving me an insight into what the optimal number of topics could look like. Assessing from the coherence scores, I intuited that the ideal number of topics would fall between 14 to 36. This intuition stemmed from the fact that, for each of the four years of the newsletter, the coherence scores increased rapidly when the topic numbers increased and reached their peaked when I set the topic number to 34. As the number of topics increased, however, the coherence scores declined gradually, indicating that topic numbers higher than 34 would likely be inaccurate. The results shown in Figure 3 gave me a good idea of the number of topics to explore the topic models.

Figure 3.
Coherence scores graph.



The figure above indicates the coherence score for the various LDA models. At around 0.49, we notice that the score plateaus, signaling that 20 topics would be the idea number of topics to use for the model.

While it may seem logical to only use the model with the highest coherence score, Mimno et al. (2011) recommend that I opted for a mixed approach to evaluate the topics. The mixture would involve the coherence score and the human intuition approaches. My decision to utilize both approaches was partly influenced by Chang et al.'s (2009) report, which indicated that held-out likelihood such as coherence and perplexity scores evaluation methods were helpful did not always produce the best results. The results from held-out likelihood methods, according to them, are in most cases unintelligible to humans. Unlike the results from Chang et al. (2009), my model's scores produced high coherence scores when the topic numbers were relatively low. Regardless, in this research, I chose to determine the best number of topics with a mixed approach. I used the human intuitive method recommended by Jockers (2014) and Chang et al. (2009), as well as the coherence score approach by Mimno et al. (2011). These methods complemented each other in their respective areas of deficiencies.

In addition to the coherence score approach, my quest for determining the optimal number of topics led me to consider the works of scholars who utilized the human intuition or judgment approach. As has been stated by some authors, including Törnberg, and Törnberg (2016), choosing the best number of topics is a qualitative process that appears more as an art than a science. In addition to the coherence scores and Törnberg and Törnberg's (2016) recommendation, I explored different numbers of topics, observing their respective outputs to ascertain those that produced the best results. I used 6, 14, 18, 20, 22, 26, 30, 40, 50, 60, and 100. The results for the six (6) and Fourteen (14) topics were too broad, encompassing more than one theme, and in some cases, the themes they generated were unintelligible. The results produced by the 50, 60, and 100 topics were too fine-grained, making them not only difficult for humans to interpret but also difficult to establish any connection to NSA's primary task – information and intelligence gathering (Debortoli, 2016). Eventually, I chose 20 as the optimal number of topics, ignoring the others.

My assessment of the topic model did not end after choosing 20 as the optimal number of topics. I explored further to determine the various topics and the newsletters the LDA model linked them. In undertaking this assessment, I manually read the newsletters to ensure that the LDA model accurately captured the content of the various newsletters. Doing this assessment helped me confirm the accuracy of the model in its assignment of topics to letters. Consider this -- the LDA model assigned *History* as the dominant topic to the April 1, 2003 newsletter. By assessing the said newsletter manually, I noticed the newsletter indeed discussed funny incidents during one of the world wars. Other similar assessments proved with convincing accuracy the model's ability to accurately assign topics to newsletters, ultimately validating that the selected number of topics was optimal.

In addition to the number of topics parameters, I tuned other parameters to obtained optimal results. Among these parameters was the number of iterations. According to Jockers (2014), the number of iteration parameters improves the model's quality when high. I, therefore, went on to increase my iteration to 500. My output showed that a model with iterations of 500 was better than those with iterations of 100, effectively confirming the recommendation in Jockers (2016).

Following my setting of the parameters and obtaining results from the model, I started to interpret the topics into themes. In undertaking this interpretation, I utilized the recommendations of Evans (2014) and Jacobi et al. (2016) and limited the topics (list of terms) to single themes or concepts. Further, I exclude those that appeared illogical. Also, complying with Evans' (2014) recommendation, I removed topics that contained more than one concept, as well as excluded those that partially consisted of some concepts. To confirm that my interpretations were accurate, I did a cross-reference of the topics with the newsletter articles that the model referenced. On average, I reviewed five (5) newsletters that linked to a topic, validating those that had high rankings among the results and presented the outcome in the Results and Discussions chapter.

Ultimately, I studied these concepts or themes to ascertain the topics the NSA covered in its newsletters, giving me an insight into some of the discussions that take place in the organization, in addition to revealing to me the latent topics of interest to the agency.

b. Does any of the topics in NSA's newsletters (SIDtoday) relate to data surveillance? What are their labels?

Data surveillance in this context refers to the deliberate and structured observation of people as a means of obtaining their data (Clarke 2003). Clarke (2003) refers to this act of collecting data about individuals or groups as dataveillance, categorizing them as processes undertaken both on an individual and mass levels. The objective of surveillance activities (data gathering and analysis) is primarily to identify patterns that could be useful for predictive purposes, or as claimed by some businesses, provide products and services more efficiently to consumers (Andrejevic & Gates, 2014; Ball et al., 2016). The NSA, however, utilizes the data in ways to achieve its SIGINT goals of spying and monitoring targeted individuals. Thus, all pursuits about collecting datasets and analyzing them to obtain insight can be referred to as data surveillance.

In light of this definition, determining topics connected to data surveillance required that the said topics meet certain expectations. A topic falls within the data surveillance spectrum if it deals with gathering, storing, and analyzing data. Therefore, any of the LDA-generated topics that concern data gathering, storing, and analyzing, came under this category.

Categorizing the topics required that I assess them, identifying and interpreting them in light of data surveillance characteristics – data gathering, storing, and analyzing. Keeping in mind these agendas, I interpreted all the topics generated by the LDA model and identified the ones that talked about data, analysis, and technologies for analyzing data. Further, I retrieved and read through newsletters that strongly connected to topics that concern data-gathering efforts, technologies for data analysis, and analysis processes to verify their accuracy. In verifying these newsletters, I noticed topics linked to data gathering discussed the data acquisition, while the data analysis topics talked about processing data. Concerning the storage and technology characteristics of data surveillance, I noticed they intertwined with the data gathering and analysis characteristics. Storage and technology characteristics are intertwined because data gathering and analysis require these features to execute their tasks. In all, the verification I undertook confirmed the model accurately identified the data surveillance topics, and the results are available in the Results and Discussions chapter.

c. What trends do we notice about NSA's data surveillance topics?

To conduct the trend analysis, I obtain the topic weights of all the topics for each of the 2121 newsletters. Topic weight is a figure that represents the degree of a topic's dominance in corpus or document or newsletter. The LDA algorithm calculates the topic weights for the newsletter. For each newsletter, the model indicates the degree of representation of all the twenty topics (20). Thus, for each newsletter, the model assigned twenty (20) weights. Likewise, for each topic, the model generated 2121 (number of newsletters) weights. Hence, I obtained 2121 topic weights for each topic and arranged them according to their publication dates (data B). Following that, I focused on the two data surveillance topics -- Topic 12 (Target Analysis) and Topic 18 (Data Collection). I then segmented the topic weights according to their respective years and calculated the average weight for each year. Thus, for each newsletter I collected the topic weights for topics 12 and 18 and organized the scores according to their respective years – 2003, 2004, 2005, 2006.

In addition to segmenting and computing the topic weight for each year, I went further to section the topic weights of each year into four (4) quarters to enable analysis quarterly. The segmentation was as follows: Quarter 1 - January to March; Quarter 2 – April to June; Quarter 3 – July to September; Quarter 4 – October to December. Thus, each year had four (4) segments.

Applying this segmentation to all the corpus yielded 16 quarters. More so, it enabled me to conduct trend analysis on a yearly and quarterly basis.

Following the quarterly segmentation, I calculated the average topic weight for each quarter. For Quarter 1, for instance, I computed the average topic weight for newsletters the agency published in January, February, and March. Taking an example from Topic 18, in the year 2004, the topic weight scores for January, February, and March were 0.032, 0.054, and 0.045, respectively. To obtain the average for this quarter, I summed all three figures and divided by three (3) to obtain 0.044. Note that the topic weights for all the twenty (20) topics sum up to 1. In essence, a percentage representation of the figures would require a multiplication of each topic weight by 100, which would indicate that Topic 18 saw a representation of 4.4% in the first quarter of 2004.

Soon after obtaining the topics weights for Topic 12 and 18, I plotted a trend graph of the topics to gain insight into their dynamics or trajectory over the four years. On the graph, each quarter's average represents a point, and there are sixteen (16) points. I assigned the topic weights on the Y-axis and the years and quarters on the X-axis.

I created a similar trend graph for the yearly trend analysis, which had only four points on the graph, as there are only four years of data for this trend analysis. As well, the arrangements of the points are in chronological order to ensure they follow along logically.

In the same vein, I utilized the approach for the quarterly trend graph for the monthly trends but without computing the topic weights quarterly. Thus, I calculated the average monthly and used the figures to generate the trend graphs.

Some challenges, however, necessitated the omission of articles from the trend analysis. An example of such a challenge is in one of the quarters. In the first quarter of 2003, there were only two articles in the corpus, making the data for that period insufficient. As a result, I removed the first quarter from the trend analysis, because it introduced errors to the study. More so, its exclusion helped to avoid using a skewed data. For instance, in computing the topic weights for Topic 12, the first quarter obtained an average of 0.136, whereas the average for all the remaining fifteen quarters was about 0.071, indicating an anomaly. Investigating the difference also revealed that the high score resulted from the first quarter having only two (2) articles. Note that these two articles are the first of SIDtoday's publication. The outcome, however, can be located in the Results and Discussions (next) chapter.

Chapter 3: Results and Discussion

In this chapter, I present the results of topic modeling on NSA's newsletters and provided descriptions of the various topics. For each of the topics, I provided labels that reflect their discourse. I have categorized some topics under broader categories that reflect their general overview. For instance, with topics like Leadership (Topic 8), Position (Topic 13), and Analyst's Skillset (Topic 17), I have sectioned them under Human Resources as a means of capturing the overall idea that permeates them. Presenting the topics in this manner made their explanation simpler. However, this presentation style did not extend to all topics, as I identified no commonalities among some. In addition to presenting all the topics the NSA discussed, I provided insights into the agency's data surveillance agendas, identifying and presenting topics that have connections to data collection and analysis efforts. For these surveillance topics, I show their yearly, quarterly, and monthly trends to describe how NSA's focus on those topics has shifted over time.

Also, these results cross-reference NSA's history and Edward Snowden's journey by confirming some information in the earlier sections of this thesis. Thus, this study verified some of the details about the NSA on matters of data gathering, customer relations, military support, and partnership. Evidently, the result from the LDA model corroborated the content of historical information about the agency. The remaining topics unraveled by the LDA models serve as additional knowledge to the existing information about the agency.

Topics Discussed by the NSA in SIDtoday

To reveal the topics in SIDtoday, I utilized the Latent Dirichlet Algorithm (LDA) to discover the main themes discussed by the agency. Table 2 shows topics that the NSA discusses in SIDtoday, which led me to identify several themes within the corpus (SIDtoday). Despite choosing twenty (20) as the ideal number of topics for this study, I omitted two topics that were unintelligible, as they added no value to this research. Regardless, all the twenty (20) topics the model generated are in Appendix A. The topics that are in Table 2, however, are those that are intelligible and relatable to the context of the SIDtoday. I concluded that these topics make sense based on the insight and the value their terms provided.

Also, I explored some of the terms of these topics in Voyant to ensure they are semantically meaningful to the overall concept prevalent in the said topics. The Lambda feature in pyLDAvis also proved helpful, allowing me to explore exclusive terms in each topic as a means of evaluating them. For instance, in a topic about Data Collection, I expect to see, almost exclusively, technical terms like antenna, signal, and satellite. Therefore, seeing these terms in these respective topics convinced me of the accuracy of the LDA model and the topics it produced. Another means of evaluation I used was to explore the corpus using these exclusive words as Key Word in Context (KWIC). KWIC assisted me not only to evaluate the topics but also to understand the context within which they occurred.

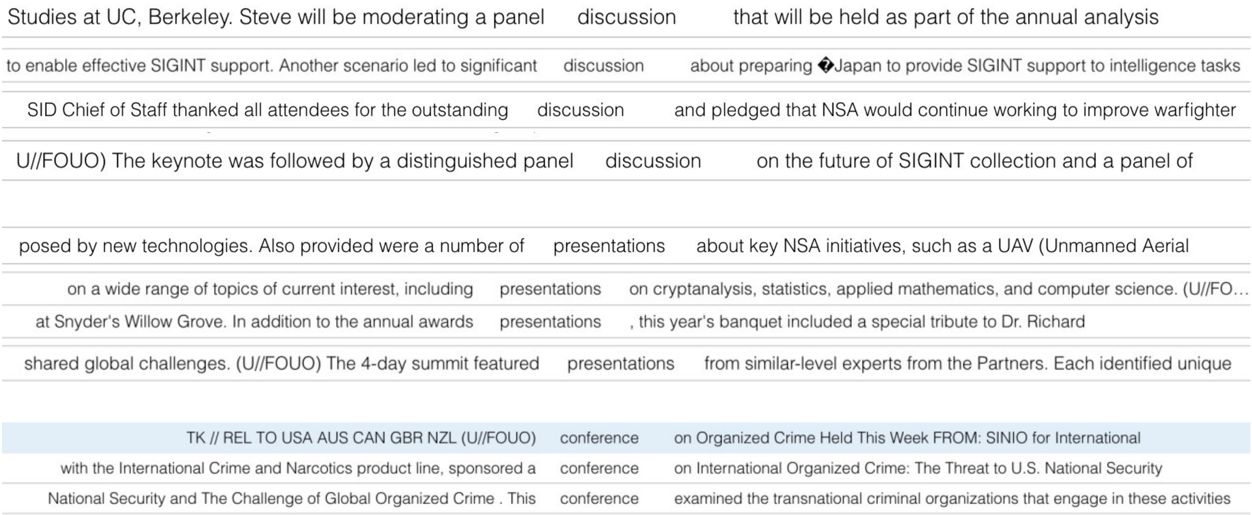
Table 2. Topics with their Labels, Representation, and Term

Dominant Topic	Topic Label	% Total Docs	Topic Desc
Topic 1	Field Work Leisure	3.77	world, life, remember, large, live, drive, city, building, day, work
Topic 2	Session or Conference	8.35	conference, session, sinio, attend, discussion, present, presentation, registration, council, development
Topic 3	Operations	4.01	operation, mission, support, center, provide, target, element, area, build, personnel
Topic 4	Military	2.92	force, military, command, north, air, korea, army, plan, joint, service
Topic 5	History	3.16	view, history, story, president, event, insider, ambassador, send, force, turn

Topic 6	Customer Service	8.11	customer, process, plan, system, requirement, management, enterprise, mission, strategy, focus
Topic 7	Information Sharing with Partners	5.47	information, partner, share, policy, security, party, foreign, office, relationship, access
Topic 8	Leadership	3.11	make, leader, good, people, technical, leadership, idea, decision, problem, important
Topic 9	Intelligence Community	4.43	intelligence, information, analysis, community, national, analyst, issue, analytic, cia, include
Topic 10	Deployment Support	4.95	iraq, team, day, baghdad, part, iraqi, deploy, support, time, hour
Topic 12	Target Analysis	8.25	target, network, analyst, analysis, tool, technology, datum, development, number, capability
Topic 13	Position	2.88	work, senior, time, office, agency, manager, position, officer, change, line
Topic 14	Writing	5.7	report, article, write, read, comment, question, reporting, word, product, find
Topic 15	National Security	4.24	terrorist, state, threat, security, government, attack, terrorism, war, group, east
Topic 16	Awards	3.58	year, award, member, organization, community, learn, annual, international, recognize, participate
Topic 17	Analysists' Skillset	4.29	language, program, training, analyst, skill, level, linguist, cryptologic, high, knowledge
Topic 18	Data Collection	5.04	collection, site, signal, system, communication, team, effort, survey, gchq, result
Topic 20	Web tools	7.31	information, web, page, user, document, cpe, contact, account, database, tool

As shown in Table 2, we get a general sense of NSA's discussions from the kinds of topics that show up in their SIDtoday newsletters. In Figure 4.1, for instance, Topic 2 (Session or Conference), which is the most prevalent topic in the corpus, focuses on discussions about information sessions, workshops, seminars, and conferences. These meetings entail informing and engaging in discussions about various issues such as briefings on enemy attacks, strategies the NSA uses, updates on their operations, ways to improve their response skills, and several others. In one of these articles published on July 5, 2005, the focus was on Russia's President – Vladimir Putin. Here, the discussions were about happenings in Russia with Putin at the center and how the country's relationship with Western countries was evolving. From Topic 2 (session or conference), we also get an insight into NSA's use of meetings as one of several modes of disseminating information in its community. It, however, is surprising that the agency gives that much attention to meetings. Regardless, it also indicates that the NSA considers the benefits of attending and organizing conferences and information sessions, including networking, expanding knowledge, presenting ideas, and exposing their staff to new information trends.

Figure 4. 1
Key Words in Context for topic 2 (Sessions and Conferences) terms depicting their use in the text.

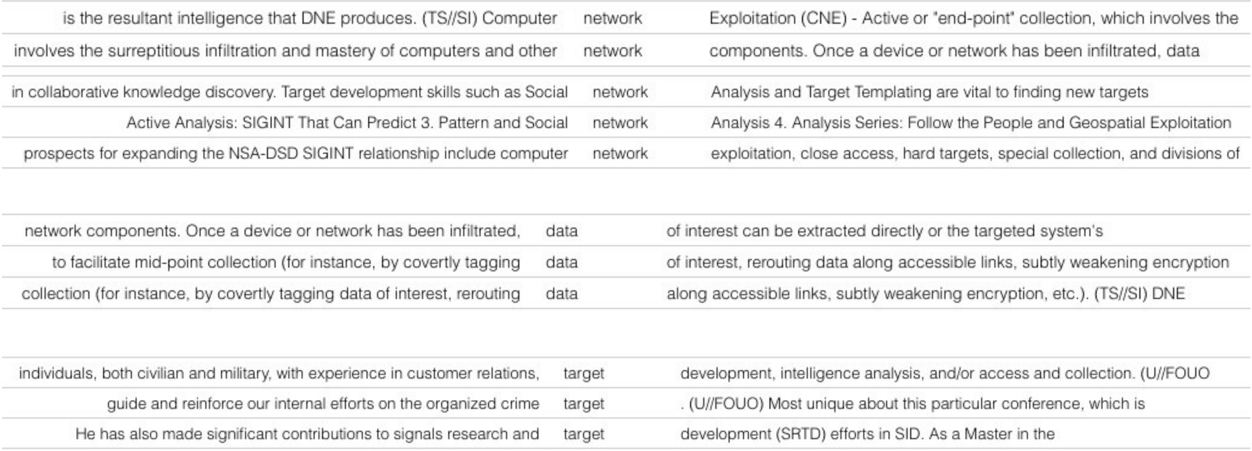


Studies at UC, Berkeley. Steve will be moderating a panel	discussion	that will be held as part of the annual analysis
to enable effective SIGINT support. Another scenario led to significant	discussion	about preparing �Japan to provide SIGINT support to intelligence tasks
SID Chief of Staff thanked all attendees for the outstanding	discussion	and pledged that NSA would continue working to improve warfighter
U//FOUO) The keynote was followed by a distinguished panel	discussion	on the future of SIGINT collection and a panel of
posed by new technologies. Also provided were a number of	presentations	about key NSA initiatives, such as a UAV (Unmanned Aerial
on a wide range of topics of current interest, including	presentations	on cryptanalysis, statistics, applied mathematics, and computer science. (U//FO…
at Snyder's Willow Grove. In addition to the annual awards	presentations	, this year's banquet included a special tribute to Dr. Richard
shared global challenges. (U//FOUO) The 4-day summit featured	presentations	from similar-level experts from the Partners. Each identified unique
TK // REL TO USA AUS CAN GBR NZL (U//FOUO)	conference	on Organized Crime Held This Week FROM: SINIO for International
with the International Crime and Narcotics product line, sponsored a	conference	on International Organized Crime: The Threat to U.S. National Security
National Security and The Challenge of Global Organized Crime . This	conference	examined the transnational criminal organizations that engage in these activities

In another vein, some of the topics reveal NSA's involvement in data collection and analytics. From the LDA results, and as shown in the KWIC in Figures 4.2 and 4.3, I noticed that Topics 12 (Target Analysis) and 18 (Data Collection) had strong links to data surveillance. Particularly, I observed that most of the newsletters that link to Topic 12 (Target Analysis) talked about data acquisition, data analysis, and tools for carrying out such goals. Many such examples are in the newsletters, and one of these is the article published on May 15, 2003 in which the agency discussed how they managed to intercept a target's phone. Related contents are in the publications for February 15, 2005 and March 17, 2005. Similar to Topic 12 (Target Analysis), I noticed a trend of data-related discussions in Topic 18 (Data Collection), as it also featured discussions that are generally concerned with data gathering. However, the main distinction between these topics is that Topic 18 is mainly concerned with data collection activities in field locations, as well as approaches for gathering data, whereas topic 12 (Target Analysis) emphasizes analyzing data (see publications for April 8, 2003; April 30, 2004; March 29, 2005; and October

10, 2006). Obviously, after collecting some form of data, an NSA analyst would need to process and make meaning of it, which is where Topic 12 (Data Analytics) comes in. To many people, however, the fact that the NSA is involved in Target Analytics and Data Collection is unsurprising, as these activities tie in well with the agency's primary agenda – intelligence gathering.

Figure 4. 2
Key Words in Context for topic 12's (Target Analysis) terms depicting their use in the newsletters.



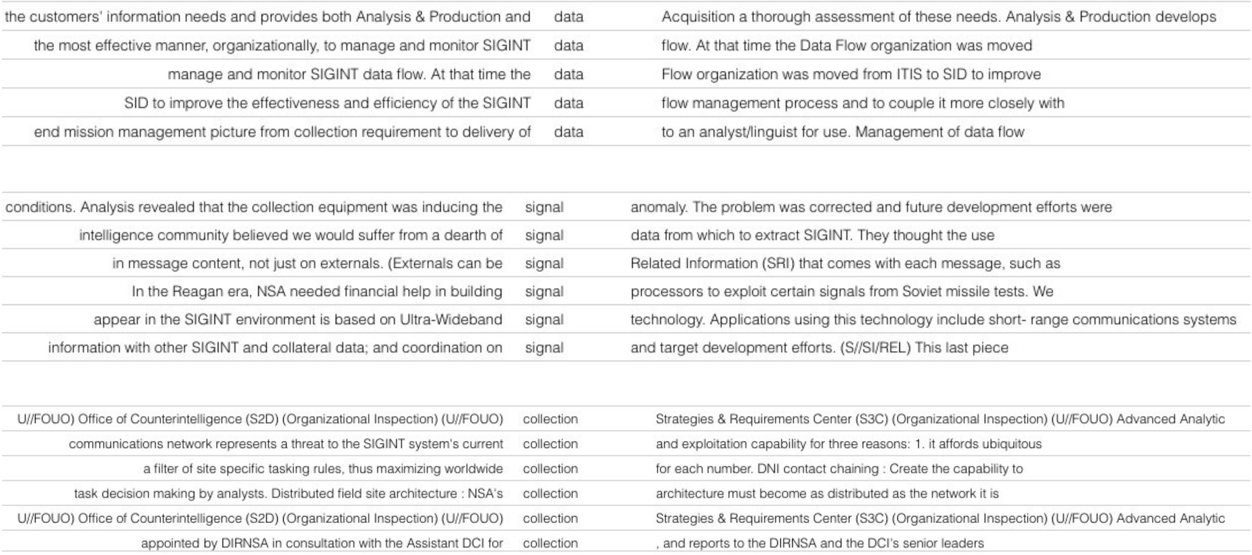
is the resultant intelligence that DNE produces. (TS//SI) Computer	network	Exploitation (CNE) - Active or "end-point" collection, which involves the
involves the surreptitious infiltration and mastery of computers and other	network	components. Once a device or network has been infiltrated, data

in collaborative knowledge discovery. Target development skills such as Social	network	Analysis and Target Templating are vital to finding new targets
Active Analysis: SIGINT That Can Predict 3. Pattern and Social	network	Analysis 4. Analysis Series: Follow the People and Geospatial Exploitation
prospects for expanding the NSA-DSD SIGINT relationship include computer	network	exploitation, close access, hard targets, special collection, and divisions of

network components. Once a device or network has been infiltrated,	data	of interest can be extracted directly or the targeted system's
to facilitate mid-point collection (for instance, by covertly tagging	data	of interest, rerouting data along accessible links, subtly weakening encryption
collection (for instance, by covertly tagging data of interest, rerouting	data	along accessible links, subtly weakening encryption, etc.). (TS//SI) DNE

individuals, both civilian and military, with experience in customer relations,	target	development, intelligence analysis, and/or access and collection. (U//FOUO
guide and reinforce our internal efforts on the organized crime	target	. (U//FOUO) Most unique about this particular conference, which is
He has also made significant contributions to signals research and	target	development (SRTD) efforts in SID. As a Master in the

Figure 4. 3.
Key Words in Context for topic 18's (Data Collection) terms depicting their use in the newsletters.



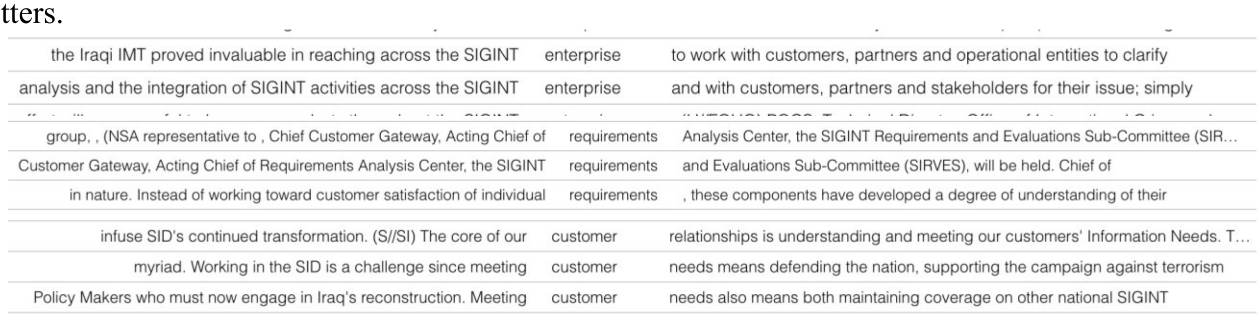
the customers' information needs and provides both Analysis & Production and	data	Acquisition a thorough assessment of these needs. Analysis & Production develops
the most effective manner, organizationally, to manage and monitor SIGINT	data	flow. At that time the Data Flow organization was moved
manage and monitor SIGINT data flow. At that time the	data	Flow organization was moved from ITIS to SID to improve
SID to improve the effectiveness and efficiency of the SIGINT	data	flow management process and to couple it more closely with
end mission management picture from collection requirement to delivery of	data	to an analyst/linguist for use. Management of data flow

conditions. Analysis revealed that the collection equipment was inducing the	signal	anomaly. The problem was corrected and future development efforts were
intelligence community believed we would suffer from a dearth of	signal	data from which to extract SIGINT. They thought the use
in message content, not just on externals. (Externals can be	signal	Related Information (SRI) that comes with each message, such as
In the Reagan era, NSA needed financial help in building	signal	processors to exploit certain signals from Soviet missile tests. We
appear in the SIGINT environment is based on Ultra-Wideband	signal	technology. Applications using this technology include short- range communications systems
information with other SIGINT and collateral data; and coordination on	signal	and target development efforts. (S//SI/REL) This last piece

U//FOUO) Office of Counterintelligence (S2D) (Organizational Inspection) (U//FOUO)	collection	Strategies & Requirements Center (S3C) (Organizational Inspection) (U//FOUO) Advanced Analytic
communications network represents a threat to the SIGINT system's current	collection	and exploitation capability for three reasons: 1. it affords ubiquitous
a filter of site specific tasking rules, thus maximizing worldwide	collection	for each number. DNI contact chaining : Create the capability to
task decision making by analysts. Distributed field site architecture : NSA's	collection	architecture must become as distributed as the network it is
U//FOUO) Office of Counterintelligence (S2D) (Organizational Inspection) (U//FOUO)	collection	Strategies & Requirements Center (S3C) (Organizational Inspection) (U//FOUO) Advanced Analytic
appointed by DIRNSA in consultation with the Assistant DCI for	collection	, and reports to the DIRNSA and the DCI's senior leaders

Although the NSA engages in all these data agendas, they are not the primary users, as their job requires them to render services to their end-users, which they mostly referred to as customers. This notion of service to customers is evident in some of their discussions. To explain, in Topic 6 (Customers), the third most prevalent topic, you would realize the NSA's customer-driven nature, as it takes into account the requirements of its clientele's products and service needs. By reading through the Customer Service dominated newsletters, you would observe discussions about products and services needed by NSA's customers, who those customers are, their changing needs, and leadership of the Customer Relations Directorate. Also, by looking closely at the newsletter you would notice NSA's list of customers include organizations in the intelligence community (Department of Defense and the military), the US Congress, and the non-Title 50 Agency (non-Intelligence Community organizations that rely on NSA's information for executive decisions (see newsletter for March 16, 2005). Besides, the newsletters reveal that within the NSA, several positions handle customer relation needs. These positions include Customer Programs Manager, Customer Relation Director (see newsletter for August 1, 2003), Customer Engagement Product Manager (see newsletter for August 4, 2004), Customer Account Manager (see newsletter for March 1, 2005). Analyzing this topic further indicated in the newsletter from October 28, 2003 (Requirements Analysis Center) that the NSA took its clients seriously, to the extent of setting up a Requirement Analysis Center (RAC) to serve them. They also had a director to manage such affairs. The KWIC in Figure 4.4 helps to corroborate this assessment. Clearly, Topic 6 reveals the NSA's customer service approach that mandates the agency to render intelligence gathering services on behalf of other agencies in the U.S.

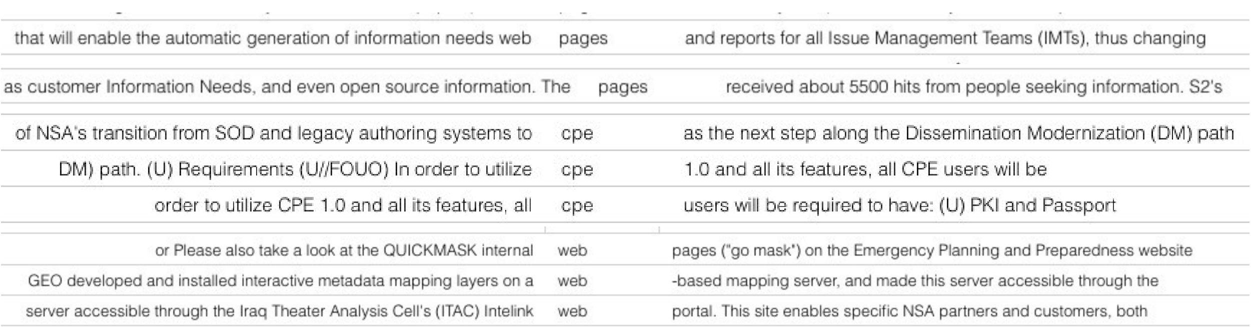
Figure 4. 4.
Key Words in Context for topic 6 (Customer Service) terms depicting their use in the newsletters.



the Iraqi IMT proved invaluable in reaching across the SIGINT	enterprise	to work with customers, partners and operational entities to clarify
analysis and the integration of SIGINT activities across the SIGINT	enterprise	and with customers, partners and stakeholders for their issue; simply
group, , (NSA representative to , Chief Customer Gateway, Acting Chief of	requirements	Analysis Center, the SIGINT Requirements and Evaluations Sub-Committee (SIR…
Customer Gateway, Acting Chief of Requirements Analysis Center, the SIGINT	requirements	and Evaluations Sub-Committee (SIRVES), will be held. Chief of
in nature. Instead of working toward customer satisfaction of individual	requirements	, these components have developed a degree of understanding of their
infuse SID's continued transformation. (S//SI) The core of our	customer	relationships is understanding and meeting our customers' Information Needs. T…
myriad. Working in the SID is a challenge since meeting	customer	needs means defending the nation, supporting the campaign against terrorism
Policy Makers who must now engage in Iraq's reconstruction. Meeting	customer	needs also means both maintaining coverage on other national SIGINT

Another topic that emerged from the LDA model contained terms related to web or online computer applications, prompting me to label it as Web Tools. This topic is Topic 20, and it indicates that the agency is heavily involved in web or online Information Technology (IT) infrastructure. Exploring further revealed that the NSA uses several IT tools to undertake its activities, of which a reasonable number of them appear to be online, and Figure 4.5 confirms this evaluation. For one, SIDtoday is one of such tools the NSA used for its work. Another example of such tools is MESSIAH, an online or web-based tool for reporting purposes (see article for May 3, 2005). In articles related to the Web Tools topic, there are also discussions about new tools such as the eDoc Exchange and modern technology platforms such as the Portal technology. Content Preparation Environment (CPE) and Public Key Infrastructure (PKI) are among the tools that have been discussed frequently in the articles. CPE was born out of a name change to avoid confusion between products (SIGINT on Demand) and tools (SIGINT on Demand). PKI, on the other hand, is an encryption tool the NSA uses to secure their transfer of information. Judging from all these tools and the several online IT tools the NSA uses, we can conclude that these tools play a significant role in NSA activities.

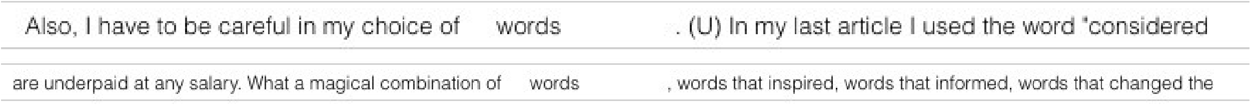
Figure 4. 5.
Key Words in Context for topic 20 (Web tools) terms depicting their use in the newsletters.



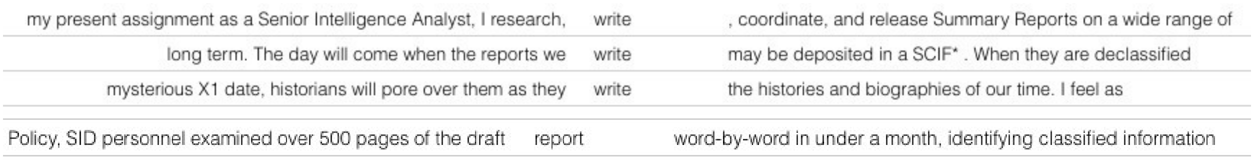
that will enable the automatic generation of information needs web	pages	and reports for all Issue Management Teams (IMTs), thus changing
as customer Information Needs, and even open source information. The	pages	received about 5500 hits from people seeking information. S2's
of NSA's transition from SOD and legacy authoring systems to	cpe	as the next step along the Dissemination Modernization (DM) path
DM) path. (U) Requirements (U//FOUO) In order to utilize	cpe	1.0 and all its features, all CPE users will be
order to utilize CPE 1.0 and all its features, all	cpe	users will be required to have: (U) PKI and Passport
or Please also take a look at the QUICKMASK internal	web	pages ("go mask") on the Emergency Planning and Preparedness website
GEO developed and installed interactive metadata mapping layers on a	web	-based mapping server, and made this server accessible through the
server accessible through the Iraq Theater Analysis Cell's (ITAC) Intelink	web	portal. This site enables specific NSA partners and customers, both

As an agency involved in obtaining and disseminating information, the NSA takes seriously many things about information. The level of importance the NSA gives to information is evident in some of the topics the LDA model revealed – Topic 14 (writing) and Topic 7 (Information Sharing with Partners). Topic 14, for instance, focused on the preparation of several documents including, reports, SIDtoday articles, and presentations. Figure 4.6 also confirms that not only does the Writing topic discuss document preparation, but it also discusses means for which writers could improve their writing through clarity and choice of vocabulary (see articles from July 18, 2005; November 15, 2005; and December 7, 2005). Whereas the writing topic strives to deal with many editorial tasks, Topic 7 (Information Sharing with Partners) captures aspects of SIDtoday's articles centered on information sharing with foreign allies and the public. The Information Sharing topic also discussed challenges the agency encounters in its information dissemination endeavors and discusses guidelines for distributing information. See a KWIC of Topic 7 in Figure 4.7 At its core, the NSA does not take for granted any issues about information. Therefore, it makes efforts to improve several aspects of its information-handling business.

Figure 4. 6.
Key Words in Context for topic 14 (Write) terms depicting their use in the newsletters.

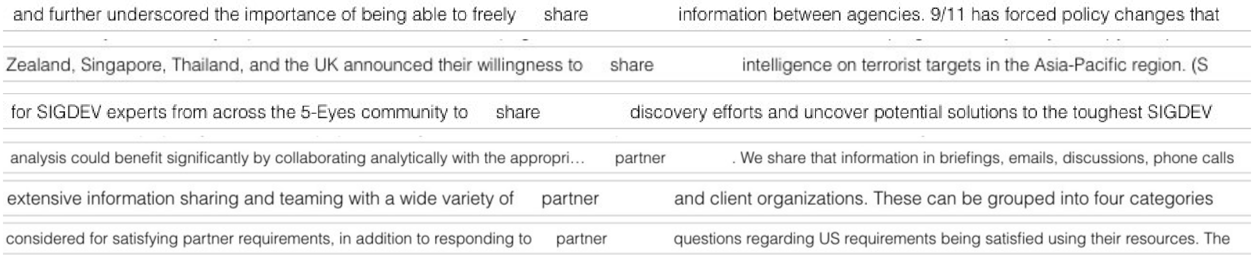


Also, I have to be careful in my choice of	words	. (U) In my last article I used the word "considered
are underpaid at any salary. What a magical combination of	words	, words that inspired, words that informed, words that changed the



my present assignment as a Senior Intelligence Analyst, I research,	write	, coordinate, and release Summary Reports on a wide range of
long term. The day will come when the reports we	write	may be deposited in a SCIF* . When they are declassified
mysterious X1 date, historians will pore over them as they	write	the histories and biographies of our time. I feel as
Policy, SID personnel examined over 500 pages of the draft	report	word-by-word in under a month, identifying classified information

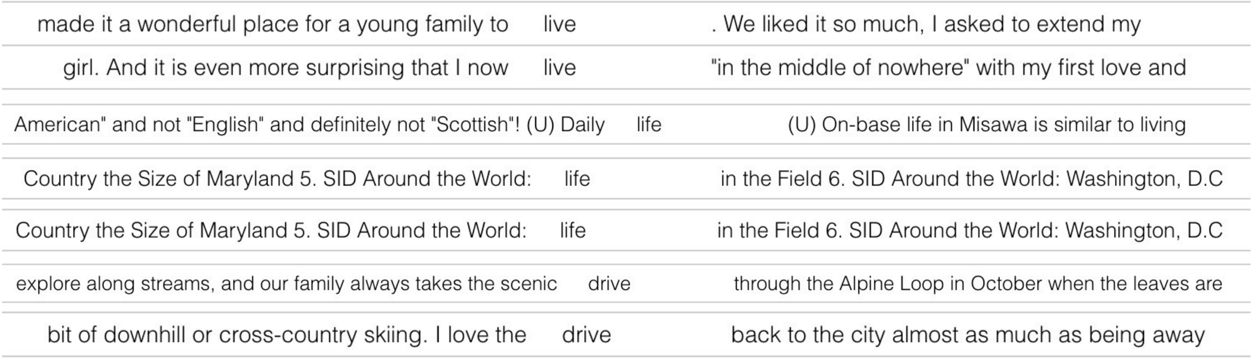
Figure 4. 7.
Key Words in Context for topic 7 (Information Sharing with Partners) terms depicting their use in the Newsletters.



and further underscored the importance of being able to freely	share	information between agencies. 9/11 has forced policy changes that
Zealand, Singapore, Thailand, and the UK announced their willingness to	share	intelligence on terrorist targets in the Asia-Pacific region. (S
for SIGDEV experts from across the 5-Eyes community to	share	discovery efforts and uncover potential solutions to the toughest SIGDEV
analysis could benefit significantly by collaborating analytically with the appropri…	partner	. We share that information in briefings, emails, discussions, phone calls
extensive information sharing and teaming with a wide variety of	partner	and client organizations. These can be grouped into four categories
considered for satisfying partner requirements, in addition to responding to	partner	questions regarding US requirements being satisfied using their resources. The

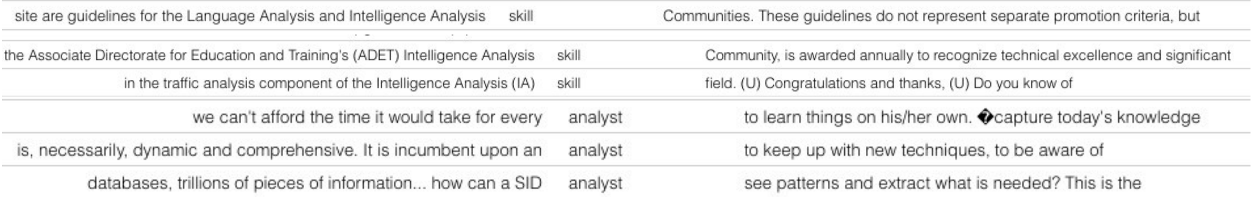
Moving on from the business of handling information, the NSA, despite its busy and mission-critical responsibilities, takes time to indulge in leisure activities, and it is evident in some of the topics the LDA model identified. These free-time activities are obvious in three topics – Field Work Leisure (Topic 1), Awards (Topic 16), and History (Topic 5). Figures 4.8, 4.9, and 4.10 also confirm these activities. In the Field Work Leisure topic, for instance, several NSA employees working abroad sent letters to their headquarters for publishing in SIDtoday activities about at their field/on-site locations (see SIDtoday newsletter for March 10, 2006). Similarly, the Awards topic (Topic 16) discussed awards, award categories, awards criterion, and recipients of various awards, while the History topic (Topic 5) included historical accounts and stories from several employees. These topics and their associated SIDtoday article point to one thing -- that the NSA consists of humans who enjoy the many pleasures life offers.

Figure 4. 8.
Key Words in Context for Topic 1 (Field Work Leisure) terms depicting their use in the Newsletters.



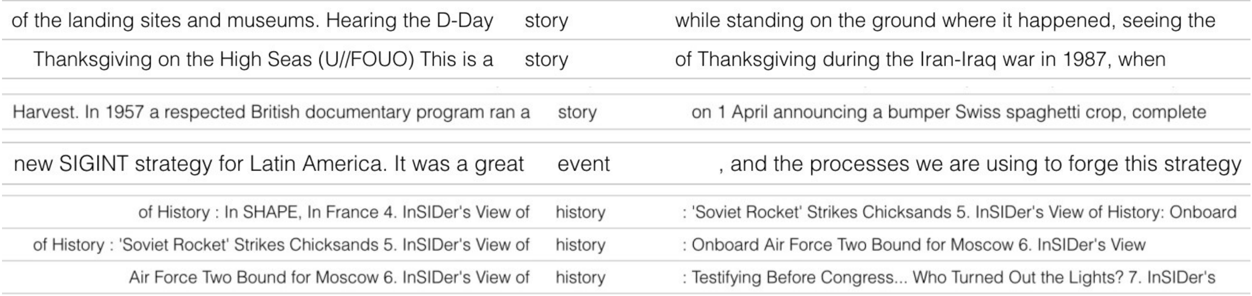
made it a wonderful place for a young family to	live	. We liked it so much, I asked to extend my
girl. And it is even more surprising that I now	live	"in the middle of nowhere" with my first love and
American" and not "English" and definitely not "Scottish"! (U) Daily	life	(U) On-base life in Misawa is similar to living
Country the Size of Maryland 5. SID Around the World:	life	in the Field 6. SID Around the World: Washington, D.C
Country the Size of Maryland 5. SID Around the World:	life	in the Field 6. SID Around the World: Washington, D.C
explore along streams, and our family always takes the scenic	drive	through the Alpine Loop in October when the leaves are
bit of downhill or cross-country skiing. I love the	drive	back to the city almost as much as being away

Figure 4. 9.
Key Words in Context for Topic 17 (Awards) terms depicting their use in the newsletters.



site are guidelines for the Language Analysis and Intelligence Analysis	skill	Communities. These guidelines do not represent separate promotion criteria, but
the Associate Directorate for Education and Training's (ADET) Intelligence Analysis	skill	Community, is awarded annually to recognize technical excellence and significant
in the traffic analysis component of the Intelligence Analysis (IA)	skill	field. (U) Congratulations and thanks, (U) Do you know of
we can't afford the time it would take for every	analyst	to learn things on his/her own. �capture today's knowledge
is, necessarily, dynamic and comprehensive. It is incumbent upon an	analyst	to keep up with new techniques, to be aware of
databases, trillions of pieces of information... how can a SID	analyst	see patterns and extract what is needed? This is the

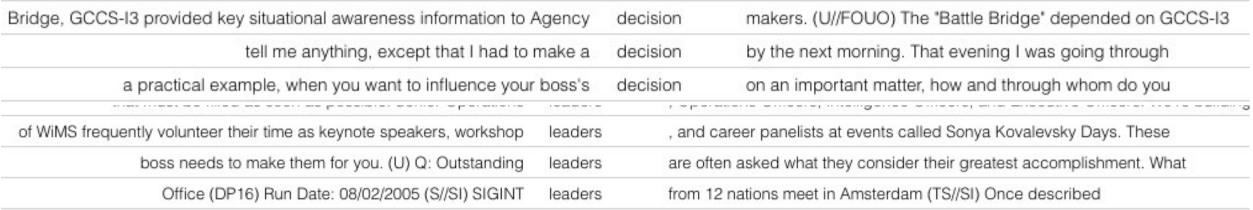
Figure 4. 10.
Key Words in Context for Topic 5 (History) terms depicting their use in the newsletter.



of the landing sites and museums. Hearing the D-Day	story	while standing on the ground where it happened, seeing the
Thanksgiving on the High Seas (U//FOUO) This is a	story	of Thanksgiving during the Iran-Iraq war in 1987, when
Harvest. In 1957 a respected British documentary program ran a	story	on 1 April announcing a bumper Swiss spaghetti crop, complete
new SIGINT strategy for Latin America. It was a great	event	, and the processes we are using to forge this strategy
of History : In SHAPE, In France 4. InSIDer's View of	history	: 'Soviet Rocket' Strikes Chicksands 5. InSIDer's View of History: Onboard
of History : 'Soviet Rocket' Strikes Chicksands 5. InSIDer's View of	history	: Onboard Air Force Two Bound for Moscow 6. InSIDer's View
Air Force Two Bound for Moscow 6. InSIDer's View of	history	: Testifying Before Congress... Who Turned Out the Lights? 7. InSIDer's

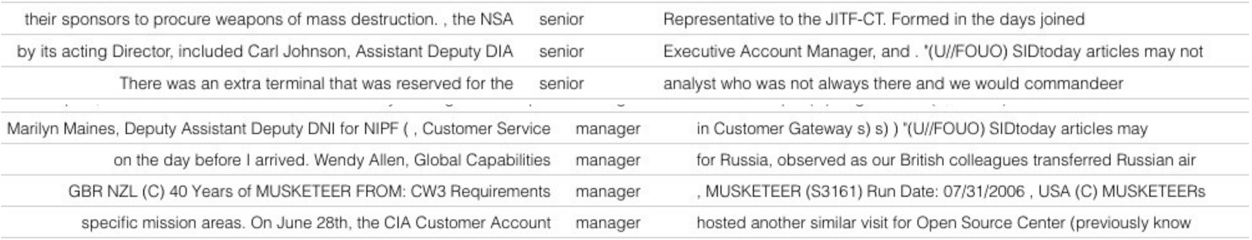
Like any other agency with humans as its employees, the NSA engages in discussions relating to its human resources. The LDA model pointed to the fact that the NSA discusses human resources topics such as Leadership (Topic 8), Position (Topic 13), and Analysts' Skillset (Topic 17). Figures 4.11, 4.12, and 4.13 explore the terms in these topics, as well as in SIDtoday's context. Concerning the Leadership topic, the discussions included issues about announcements from leaders, changes in leadership at various positions, and meetings held by NSA's leaders. An example of this discussion is in SIDtoday's article for October 14, 2003, where Major General Richard Quirk III assumed a new position as the Signal Intelligence Director. Like the Leadership topics, the Position topic (Topic 13) engaged in discussions about job offerings and responsibilities of various roles. One of the articles linked to Topic 13, for instance, referred to a Liaison Office position in Guantanamo Bay that the agency needed someone to fill. The Analyst Skillset topic (Topic 17), like the other human-resource-related topics, also makes its offerings to NSA's employees. In particular, the Analyst Skillset topic (Topic 17) relates to ways the NSA assists its signals intelligence analysts' workers (mainly analysts) to improve their skillsets. It also makes pronouncements about deficiencies in the NSA's employees' skillset and what they are doing about it (see SIDtoday article for June 29, 2004). From these LDA extracted topics, we can conclude that the NSA takes its human resource affairs seriously.

Figure 4. 11.
Key Words in Context for topic 8 (Leadership) terms depicting their use in the newsletter.

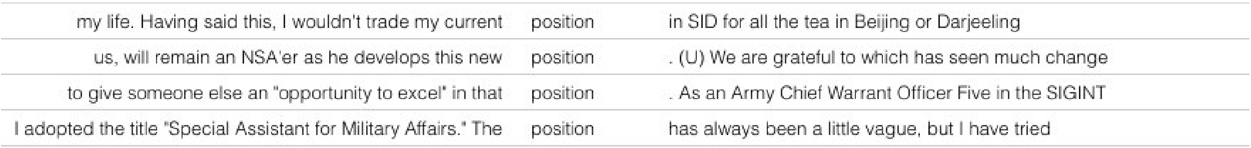


Bridge, GCCS-I3 provided key situational awareness information to Agency	decision	makers. (U//FOUO) The "Battle Bridge" depended on GCCS-I3
tell me anything, except that I had to make a	decision	by the next morning. That evening I was going through
a practical example, when you want to influence your boss's	decision	on an important matter, how and through whom do you
of WiMS frequently volunteer their time as keynote speakers, workshop	leaders	, and career panelists at events called Sonya Kovalevsky Days. These
boss needs to make them for you. (U) Q: Outstanding	leaders	are often asked what they consider their greatest accomplishment. What
Office (DP16) Run Date: 08/02/2005 (S//SI) SIGINT	leaders	from 12 nations meet in Amsterdam (TS//SI) Once described

Figure 4. 12.
Key Words in Context for topic 13 (Position) terms depicting their use in the newsletter.

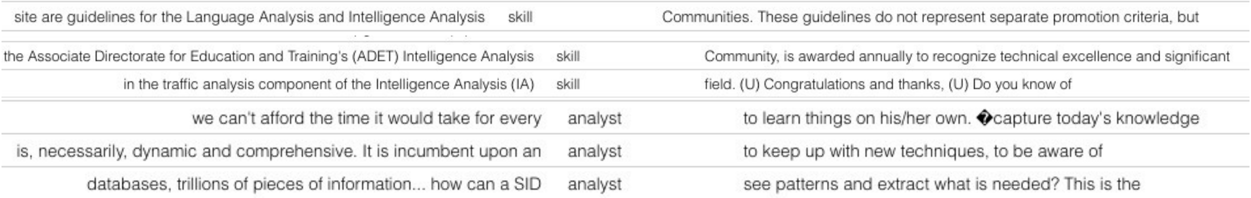


their sponsors to procure weapons of mass destruction. , the NSA	senior	Representative to the JITF-CT. Formed in the days joined
by its acting Director, included Carl Johnson, Assistant Deputy DIA	senior	Executive Account Manager, and . "(U//FOUO) SIDtoday articles may not
There was an extra terminal that was reserved for the	senior	analyst who was not always there and we would commandeer
Marilyn Maines, Deputy Assistant Deputy DNI for NIPF (, Customer Service	manager	in Customer Gateway s) s)) "(U//FOUO) SIDtoday articles may
on the day before I arrived. Wendy Allen, Global Capabilities	manager	for Russia, observed as our British colleagues transferred Russian air
GBR NZL (C) 40 Years of MUSKETEER FROM: CW3 Requirements	manager	, MUSKETEER (S3161) Run Date: 07/31/2006 , USA (C) MUSKETEERs
specific mission areas. On June 28th, the CIA Customer Account	manager	hosted another similar visit for Open Source Center (previously know



my life. Having said this, I wouldn't trade my current	position	in SID for all the tea in Beijing or Darjeeling
us, will remain an NSA'er as he develops this new	position	. (U) We are grateful to which has seen much change
to give someone else an "opportunity to excel" in that	position	. As an Army Chief Warrant Officer Five in the SIGINT
I adopted the title "Special Assistant for Military Affairs." The	position	has always been a little vague, but I have tried

Figure 4. 13.
Key Words in Context for topic 17 (Analyst Skillset) terms depicting their use in the newsletters.

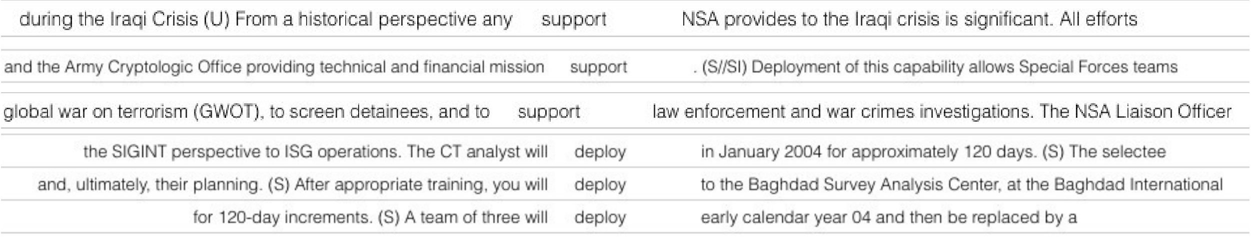


site are guidelines for the Language Analysis and Intelligence Analysis	skill	Communities. These guidelines do not represent separate promotion criteria, but
the Associate Directorate for Education and Training's (ADET) Intelligence Analysis	skill	Community, is awarded annually to recognize technical excellence and significant
in the traffic analysis component of the Intelligence Analysis (IA)	skill	field. (U) Congratulations and thanks, (U) Do you know of
we can't afford the time it would take for every	analyst	to learn things on his/her own. �capture today's knowledge
is, necessarily, dynamic and comprehensive. It is incumbent upon an	analyst	to keep up with new techniques, to be aware of
databases, trillions of pieces of information... how can a SID	analyst	see patterns and extract what is needed? This is the

As far as support goes, the NSA offers support services to various organizations, among which the military services are its main customers. From the corpus, the LDA model extracted topics that indicated that the NSA offers several support services to the military, and in some cases, deploys its employees to military sites. More so, some military persons work as analysts with the NSA. Figures 4.14 and 4.15 reflect how terms related to Topic 10 (Deployment Support) and Topic 4 (Military) are used in SIDtoday's context. I noticed it through articles related to Topic 10 (Deployment Support) and Topic 4 (Military). Particularly for Topic 10 (Deployment Support), the agency made deployments of its staff to locations that include Iraq, Baghdad, Kuwait, Afghanistan, Haiti, and Kabul to help with the unrest in those locations. Also, Topic 10 talks about the search for volunteers who would be dispatched to the deployment locations, the responsibilities of the deployed, experiences of various deployed employees, and the stories of hostages (see SIDtoday articles for October 14, 2005, and April 1, 2003). Topic 4 (Military), on the other hand, focused on the military capability of enemies (North Korea, Russia, etc.) of the US, feedback on NSA's support to the military, functions of some military centers (e.g., Joint War Analysis Center), the setup of new centers, and basic military codes (see SIDtoday articles for January 16, 2004; November 24, 2004; July 19, 2006; and July 24, 2006). Undoubtedly, these topics demonstrate the

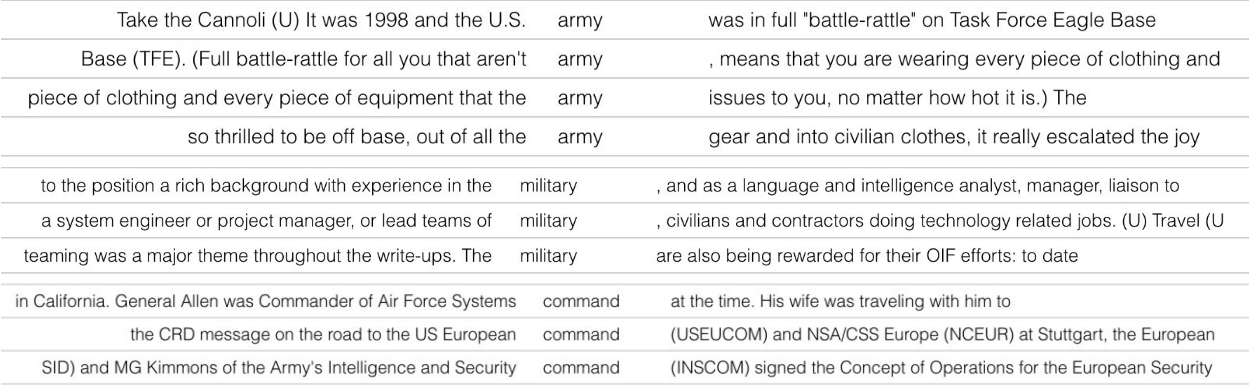
NSA’s discussion of the military services and their commitment to working with them to achieve their goals.

Figure 4. 14.
Key Words in Context for topic 10 (Deployment Support) terms depicting their use in the newsletters.



during the Iraqi Crisis (U) From a historical perspective any	support	NSA provides to the Iraqi crisis is significant. All efforts
and the Army Cryptologic Office providing technical and financial mission	support	. (S//SI) Deployment of this capability allows Special Forces teams
global war on terrorism (GWOT), to screen detainees, and to	support	law enforcement and war crimes investigations. The NSA Liaison Officer
the SIGINT perspective to ISG operations. The CT analyst will	deploy	in January 2004 for approximately 120 days. (S) The selectee
and, ultimately, their planning. (S) After appropriate training, you will	deploy	to the Baghdad Survey Analysis Center, at the Baghdad International
for 120-day increments. (S) A team of three will	deploy	early calendar year 04 and then be replaced by a

Figure 4. 15.
Key Words in Context for topic 4 (Military) terms depicting their use in the newsletters.

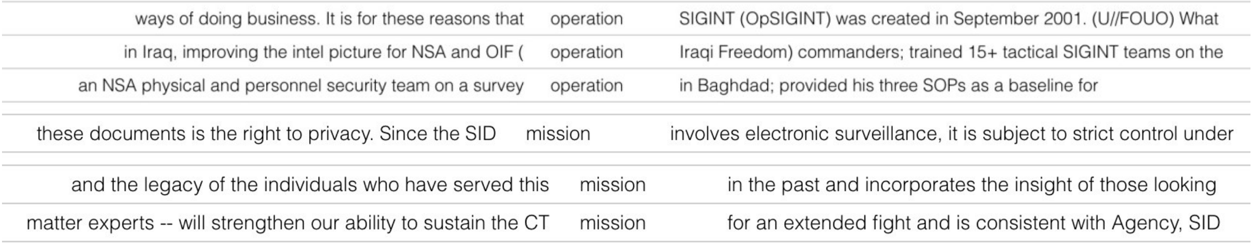


Take the Cannoli (U) It was 1998 and the U.S.	army	was in full "battle-rattle" on Task Force Eagle Base
Base (TFE). (Full battle-rattle for all you that aren't	army	, means that you are wearing every piece of clothing and
piece of clothing and every piece of equipment that the	army	issues to you, no matter how hot it is.) The
so thrilled to be off base, out of all the	army	gear and into civilian clothes, it really escalated the joy
to the position a rich background with experience in the	military	, and as a language and intelligence analyst, manager, liaison to
a system engineer or project manager, or lead teams of	military	, civilians and contractors doing technology related jobs. (U) Travel (U
teaming was a major theme throughout the write-ups. The	military	are also being rewarded for their OIF efforts: to date
in California. General Allen was Commander of Air Force Systems	command	at the time. His wife was traveling with him to
the CRD message on the road to the US European	command	(USEUCOM) and NSA/CSS Europe (NCEUR) at Stuttgart, the European
SID) and MG Kimmons of the Army's Intelligence and Security	command	(INSCOM) signed the Concept of Operations for the European Security

Other topics of interest that the LDA model revealed are Topic 3 (Operations), Topic 9 (Intelligence Community), and Topic 15 (National Security). For the Operations topic, the focus was on NSA’s operations or missions all over the world, and it gives accounts of their success and failures, as well as provides information about the locations of the agency’s operation centers and the activities going on there (see SIDtoday article for February 27, 2006). In Topic 9, the NSA discussed issues about the Intelligence Community and subjects that need address by its members. In the SIDtoday article for May 20, 2003, for instance, the NSA center for China and Korea organized an inter-agency meeting to inform other community members about an outbreak of

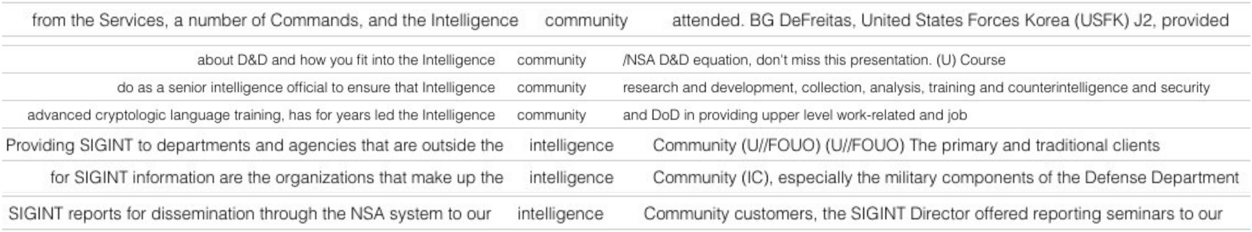
Severe Acute Respiratory Syndrome (SARS). In some respects, Topic 15 (National Security) also focused on collaborations between members of the Intelligence community and focuses on the overall national intelligence needs of the USA and its partners. The KWICs below exhibit how terms associated with these three topics are used in the newsletters. These topics give insights into the array of unique issues that are prevalent within the NSA.

Figure 4. 16.
Key Words in Context for topic 3 (Operations) terms depicting their use in the newsletters.



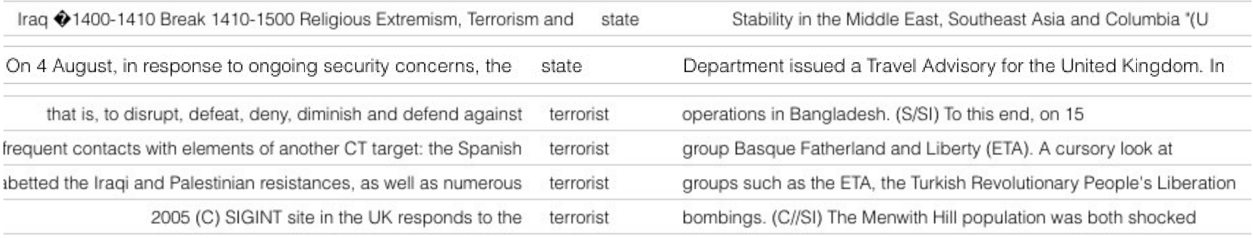
ways of doing business. It is for these reasons that	operation	SIGINT (OpSIGINT) was created in September 2001. (U//FOUO) What
in Iraq, improving the intel picture for NSA and OIF (	operation	Iraqi Freedom) commanders; trained 15+ tactical SIGINT teams on the
an NSA physical and personnel security team on a survey	operation	in Baghdad; provided his three SOPs as a baseline for
these documents is the right to privacy. Since the SID	mission	involves electronic surveillance, it is subject to strict control under
and the legacy of the individuals who have served this	mission	in the past and incorporates the insight of those looking
matter experts -- will strengthen our ability to sustain the CT	mission	for an extended fight and is consistent with Agency, SID

Figure 4. 17.
Key Words in Context for topic 9 (Intelligence Community) terms depicting their use in the newsletters.

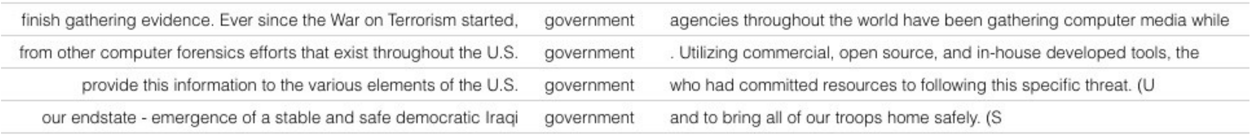


from the Services, a number of Commands, and the Intelligence	community	attended. BG DeFreitas, United States Forces Korea (USFK) J2, provided
about D&D and how you fit into the Intelligence	community	/NSA D&D equation, don't miss this presentation. (U) Course
do as a senior intelligence official to ensure that Intelligence	community	research and development, collection, analysis, training and counterintelligence and security
advanced cryptologic language training, has for years led the Intelligence	community	and DoD in providing upper level work-related and job
Providing SIGINT to departments and agencies that are outside the	intelligence	Community (U//FOUO) (U//FOUO) The primary and traditional clients
for SIGINT information are the organizations that make up the	intelligence	Community (IC), especially the military components of the Defense Department
SIGINT reports for dissemination through the NSA system to our	intelligence	Community customers, the SIGINT Director offered reporting seminars to our

Figure 4. 18.
Key Words in Context for topic 15 (National Security) terms depicting their use in the newsletters.



Iraq �1400-1410 Break 1410-1500 Religious Extremism, Terrorism and	state	Stability in the Middle East, Southeast Asia and Columbia "(U
On 4 August, in response to ongoing security concerns, the	state	Department issued a Travel Advisory for the United Kingdom. In
that is, to disrupt, defeat, deny, diminish and defend against	terrorist	operations in Bangladesh. (S/SI) To this end, on 15
frequent contacts with elements of another CT target: the Spanish	terrorist	group Basque Fatherland and Liberty (ETA). A cursory look at
ibetted the Iraqi and Palestinian resistances, as well as numerous	terrorist	groups such as the ETA, the Turkish Revolutionary People's Liberation
2005 (C) SIGINT site in the UK responds to the	terrorist	bombings. (C//SI) The Menwith Hill population was both shocked



finish gathering evidence. Ever since the War on Terrorism started,	government	agencies throughout the world have been gathering computer media while
from other computer forensics efforts that exist throughout the U.S.	government	. Utilizing commercial, open source, and in-house developed tools, the
provide this information to the various elements of the U.S.	government	who had committed resources to following this specific threat. (U
our endstate - emergence of a stable and safe democratic Iraqi	government	and to bring all of our troops home safely. (S

In a nutshell, these eighteen (18) topics do not only inform us about the diverse interests and discussions that go on within the NSA and the intelligence community, but they enlighten us about the meticulous structures they have in place to ensure success in their work. Yet, they give us insights into the mundaneness of the agency, as well.

Topics that are related to Data Surveillance

Out of the twenty topics generated by the LDA algorithm, I observed that two (2) had a strong connection with data and surveillance activities conducted by the agency. These are the topics – Topic 12 (Target Analysis) and Topic 18 (Data Collection). Although both topics fall under the concept of data, each focuses on a different specific aspect of the entire data surveillance spectrum.

Speaking of data and surveillance, Topic 12's (Target Analysis) relation to data surveillance is observable from the top terms in the topic. Analyzing and exploring Topic 12's top-ten terms (target, network, analyst, analysis, tool, technology, datum, development, number, and capability) with the KWIC tool (as shown in Figure 4.2) reveals NSA's activities involved in data surveillance. In SIDtoday, for example, a term like "network" refers to the mode of transmission NSA's targets use for communication, mostly referring to Internet Service Providers (ISPs) and computer networks, as well as the communication tools the agency's employees use to accomplish their goals. In a July 16, 2003 newsletter publication relating to networks, for instance, the author states that:

> Once a device or network has been infiltrated, data of interest can be extracted directly or the targeted system's operation can be modified to facilitate mid-point collection (for instance, by covertly tagging data of interest, rerouting data along accessible links, subtly weakening encryption, etc.)

This statement confirms that “network” in this context refers to a computer network. Also, with an occurrence of 2028 in the corpus, the term “target” for instance, after selecting at random twenty (20) of its instances and exploring them in the articles, indicated that the NSA had regularly observed locations, institutions, and persons of interest to its agenda or those of its clients. Further to its analytic goals, the agency provides for its employees sophisticated tools to undertake activities such as identifying, monitoring, and obtaining information from their targets (See article for March 31, 2003). On the other hand, the terms “analyst” and “analysis” refer to NSA’s employees who extract meaningful facts from data obtained from targets and the processes for doing so, respectively, whereas “tools,” “technology,” “development,” and “capacity” point to the facilities analyst utilize to achieve this goal on “datum” (singular for data). With its capacity to create a correlation between terms that frequently occur together, the LDA model has indicated that, in many ways, Topic 12 discusses data surveillance.

Another topic that falls under the data surveillance spectrum is Topic 18 (Data Collection). Exploring Topic 18’s terms (collection, site, signal, system, communication, team, effort, survey, GCHQ, result) with KWIC (as shown in Figure 4.3) provided a better understanding of how the NSA and its partners handled data collections. A term like “collection,” with a frequency of 1207, revealed the agency’s discussions about data collection assignments, the units responsible for those collections, their adaptation efforts to evolving changes, and their cooperation with their partners on data collection. The terms “site,” “system,” and “signal” point respectively to the overall field locations where NSA has its systems for collection data, whiles “signal” refers to the transmitted elements the sites gather for analysis. A term like “GCHQ” is an acronym for the UK’s Government Communication Headquarters, a partner of the NSA's, and has often collaborated with the agency on many signal intelligence activities, including data collection (see SIDtoday article for July 2, 2004). The terms “survey” and “result” point to one of the methods employed by the NSA in their data collection activities, as well as the outcome of that approach. From SIDtoday’s July 9, 2003 article, for instance, there is evidence of the agency’s discussion of a survey requirement in Afghanistan to establish the usage of communication channels. By exploring the terms in Topic 18, I found enough evidence to confirm the NSA’s involvement in data collection, a huge component in the data surveillance spectrum.

Exploring the corpus revealed that data surveillance is convincingly a big part of NSA’s signal intelligence discussion agendas, and some of the topics that the LDA model generated from the corpus confirmed this hypothesis. For an agency whose responsibility is to meet its customers' signal intelligence needs, it is not surprising to see Topic 12 (Data Analysis) and Topic 18 (Data Collection) in their newsletters. Nonetheless, investigating their newsletter was essential for obtaining confirmation of this kind. Also, the suspicion that the agency is involved in data surveillance became evident after reviewing the top ten terms in the topics related to data surveillance. Likewise, another review of the terms using KWIC validated NSA’s data surveillance activities.

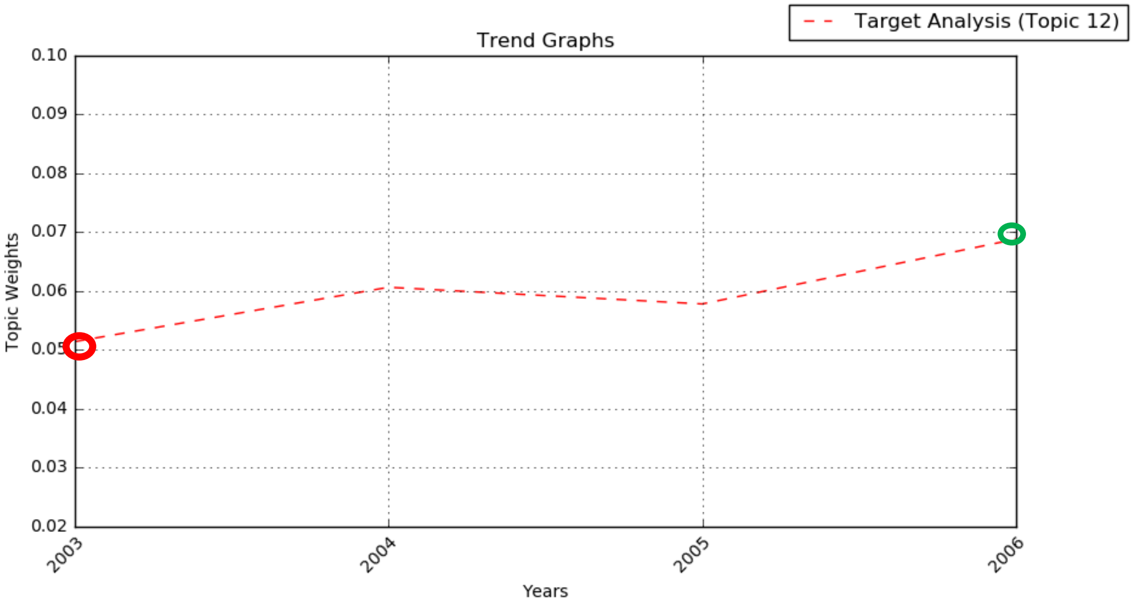
Dynamics in Data Surveillance

In analyzing the dynamics that permeate data surveillance within the SIDtoday, I utilize a trend graph to understand the changes evident in Topic 12 (Data Collection) and Topic 18 (Target Analysis). Dynamics in this context refers to the changes that data surveillance-related have undergone in SIDtoday. To understand these dynamics, I observed the trend of topics in the corpus, establishing whether data surveillance discussions within SIDtoday have declined, increased, or remained steady. Besides, this approach helped to identify periods within the respective years when the NSA focused more on data surveillance.

Topic 12

Observing the trend graph of Topic 12 (Target Analysis) reveals some interesting patterns in how the NSA handles discussions about their targets. I show these patterns in Figures 5.1, 5.1.1, 6.1, and 6.2. From the graph in Figure 5.1, which represents the topic's weights according to their respective years, we can see a consistent upward trajectory of Topic 12 from 2003 to 2006. This trajectory indicates a gradual and unfluctuating growth of the NSA's interest in its targets every year.

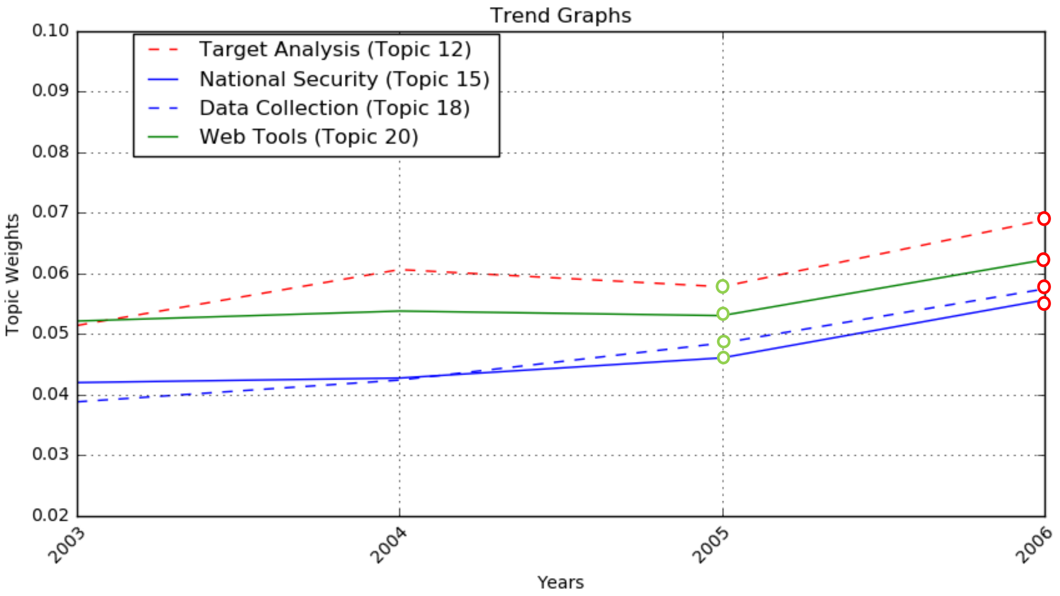
Figure 5. 1
Trend Graph of Topic 12 (Target Analysis) Plotted Annually.



A historical trend of Topic 12's trajectory of 0.051 (marked in red) in 2003 to 0.069 (marked in green) in 2006

In explaining factors that influenced Topic 12's trajectory, I noticed that concepts such as new threats and new technologies were responsible for the upward trajectory. Thus, the increase in threats and improvements in technologies for the U.S. and its allies were possible reasons for Topic 12's increment. For instance, if there is a war or terrorist attack, the U.S. government would endeavor to prepare all necessities to increase its chance of victory. The NSA's SIGINT products would be among the vital resources for such undertakings. Therefore, in such situations, the NSA would strive to meet the intelligence needs of the U.S., thereby increasing its target analysis of threats. As such, we are likely to see some similarities between Topic 12's trajectory that relate to new threats and technologies – Topic 15 (National Security) and Topic 20 (Web tools). Also, I included Topic 18's trend, as issues regarding increased data collection would likely impact target analysis endeavors. In Figure 5.1.1, despite the nuanced difference, we see significant similarities in the yearly trajectories for Topics 12, 15, 18, and 20. For instance, in 2006, we noticed that all four (4) topics surged up, indicating a possible correlation between them.

Figure 5.1. 1
Yearly Trend Graph for Topic 12, 15, 18, and 20.

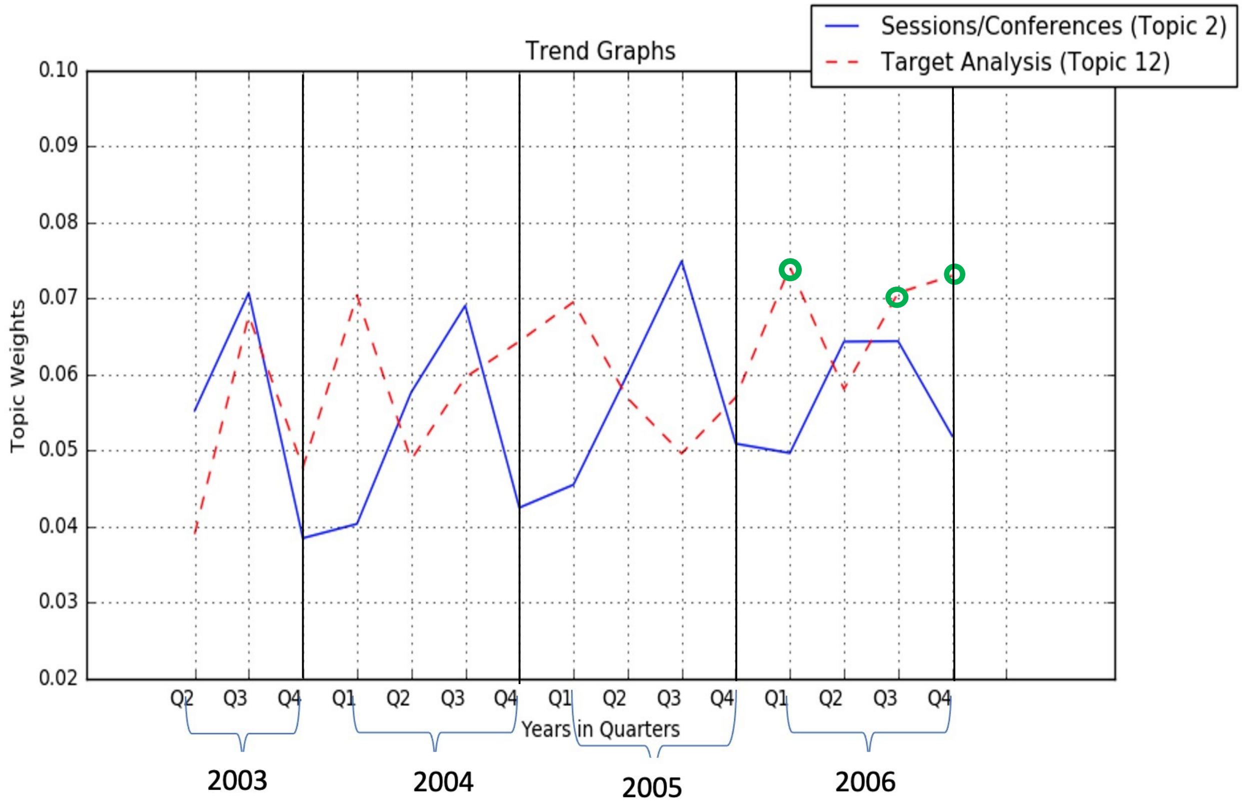


In the above figure, we notice a similar trend in all the topics, particularly in 2006, to suggest a correlation. The green markers in 2005 indicate the point where the topics made a significant upward turn towards 2006. The red marks their respective highs in 2006.

In addition to the commonalities revealed by the trend graphs in Figure 5.1.1, there are pointers in SIDtoday that confirm that Topic 12's rise was due in part to influences from Topics 15, 18, and 20. In SIDtoday's release for January 3, 2005, for instance, the author revealed that technological advancement (Web Tools related) led to an improvement in the NSA's target analysis (Target Analysis related) activities by stating that "a new reporting tool is now available that will help both the analyst and mission manager make more informed decisions on target development and tasking (p. 1)." Also, on December 21, 2006, SIDtoday's writer made a somewhat indirect connection between national security and target analysis by recounting that "it has been a challenging year, with the campaigns in Iraq and Afghanistan, the larger Global War on Terrorism, and a host of other issues all demanding top-notch SIGINT support (p. 1).” Despite the lack of an obvious connection, we can infer that an increase in demand for SIGINT would mean analysts at the NSA needed to undertake more target analytics works, which ultimately reflects in discussions engaged in SIDtoday. These similarities in Topics 12, 15, 18, and 20 and the intertwined connections between them confirm the impact they extend onto one another, hence Topics 15, 18, and 20’s influence on Topic 12.

From Figure 6.1, however, we get an in-depth view of Topic 12’s trends in a quarterly representation. Here, we notice that the NSA’s discussions centered on Topic 12 (Target Analysis) were considerably high in the first quarter (Q1) when compared to all the other years with an average of 0.07 (7%), which is also beyond that of all the other quarter. NSA’s consistent expression of interest in Target Analysis-related topics in Q1 (January, February, and March) led to further exploration of related newsletters to obtain a better understanding of the trends. Upon reviewing (or close reading) the related newsletters, I noticed that the spike in Topic 12 in the first quarters (Q1) resulted from the NSA’s increased discussion of new tools for analysis targets purposes in Q1.

Figure 6. 1.
Quarterly Trend Graph of Topic 2 and 12.



A meandering quarterly display of Topic 2 (in blue line) and Topic 12 (in red line) from 2003 to 2006. We notice marginal rises (marked in green) in Topic 12 in 2006 compared to the previous years.

To confirm this finding, I did a frequency count of terms that entail the concept of new analytic tools. Here, Table 3 shows a simple frequency count of the terms “new,” “tool,” and “tools.” Table 1 indicates that the first quarter had a high total count of the terms for all the four quarters, confirming the pattern that the NSA engages more in discussions about new target analysis tools in the first quarter of most years. I went further to search for the terms “new year” and “new tools” in the corpus to ascertain any ambiguity about “new" in this context. A simple frequency count revealed that “new tool” occurred twenty-nine (29) times, whereas “new year” occurred twenty-two times (22). For “new year,’ when I focused on the frequency counts on January, February, and March, I recorded ten (10) occurrences, which is insignificant to indicate any ambiguation in the analysis. As such, there was no doubt that the first quarter focused on new target analysis. Clearly, the spikes in the first quarter of each year are enough evidence that the NSA continuously renewed preparedness and posture to tackle its tasks of analyzing targets at the beginning of each year.

Table 3. Frequency Count of these Terms: “new,” “tool,” and “tools.”

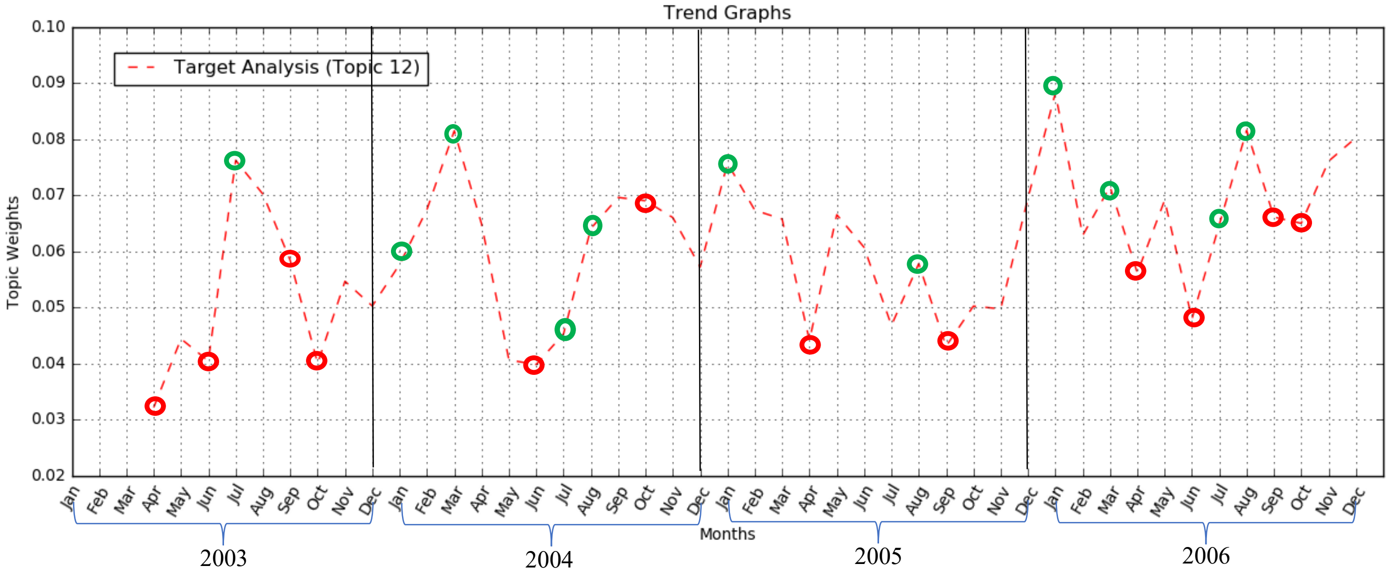
Terms	**Q1**	**Q2**	**Q3**	**Q4**
New	62	44	70	45
Tool	76	22	24	37
Tools	78	29	46	35
Total Frequency Count	**216**	**95**	**140**	**117**

Contrary to NSA’s increased interest in Topic 12 (Target Analysis) during the first quarters of the various years, in the second quarter (Q2), we see a decline in the topic, as revealed by the trend graph in Figure 6.1. Analyzing this decline revealed that during the second quarter (Q2) of 2003, the NSA engaged less in discussions about Topic 12 but paid increased attention to Web Tools (Topic 20). Like in 2003, the second quarter (Q2) of the subsequent years saw dips in Topic 2 (Sessions/Conferences), with reasons similar to Q1’s. In 2004, for instance, Topic 2 (Sessions/Conferences) received most of the attention, and so did 2005. Also, in the Q2 of 2006, Topic 12 (Target Analysis) declined but with a nuance. 2006’s Q2 received more attention than the Q2s of the other years. More so, despite its dip, the attention Topic 12 received in Q2 of 2006 surpassed all other topics except Topic 15’s (National Security), Topic 2’s (Sessions/Conferences), and Topic 12's (Data Collection). This occurrence is mainly because the focus on Topic 12 has gradually increased over the years despite the frequent plummets it experiences in between quarters. Although the number of published newsletters for the various years remained the same, there was a consistent increase in Topic 12 (Target Analysis) over the years. This trend resulted in the later years receiving more Target Analysis discussions attention than the former years. In all, the trends reveal that although the NSA's focus on Topic 12 reduces in some months, their overall interest did not subside during the four (4) years period of this data.

Unlike the nearly consistent trajectory of Topic 12 in the second quarter of the various years, the trend in the third quarter for the said topic is slightly different. Topic 12 (Target Analysis) had an undulating trend in the third quarter (Q3) such that all the years, except 2005, surged up. While Topic 12 received topic weights of 0.068, 0.06, and 0.07 for the third quarters in 2003, 2004, and 2006, respectively, 2005's plummeted to 0.05, making the third quarter of the year 2005 an exception when compared to the rest. Assessing the trend of Q3's dip in 2005 revealed that the NSA focused more on conference-related (Topic 2) issues during those times of the year.

For the fourth quarter (Q4) of Topic 12, it is noticeable from Figure 6.1 that the fourth quarter showed an undulating pattern, making it similar to the third quarter. We can also tell that, except for 2003 where a dip occurred, the topic weights of Topic 12 remained somewhat consistent in 2004 and 2005 but soared in 2006. While this pattern indicates that the NSA has no particular patterns for discussing Topic 12 (Target Analysis) related matters in the fourth quarter, it reveals the agency's persistent interest in the topic. Thus, observing the rise in three out of four quarters suggests that the NSA prioritized its target analysis endeavors throughout the year.

Figure 6. 2.
Monthly Trend Graph for Topic 12 (Target Analysis).



Red circles indicate months where Topic 12 increased, green indicate months when the topic decreased. The black vertical lines (not the grids) mark the end of the calendar years for our sample period. This trend shows that, monthly, NSA's attention to Topic 12 does not follow any pattern.

Monthly, we get a perspective on the NSA focus on Topic 12's (Target Analysis) discussions. The graph in Figure 6.2 reveals a trend similar to those in Figures 5.1, 5.1.1, and 6.1, respectively. We notice that Figure 6.2 shows undulating patterns in the way the NSA discusses target analysis topics monthly. Yet, the focus the agency had for the topic remained somewhat similar throughout the years. For some specific months, however, we notice some revealing patterns. For instance, in January, we see more rises than drops throughout the years. A similar trend applies to March, July, and August. For April, June, September, and October, we see more frequent drops than rises, indicating that the NSA's focus on the topic decreased during those months. From this perspective, we get an indication of the NSA's irregular attention to target analysis topics.

From Figure 6.2, we also notice that within the different years, some months recorded their respective maximum drops or rises. For instance, in July, March, January, and January for 2003, 2004, 2005, 2006, respectively, we saw sudden peaks, signaling the NSA's increased attention to target analysis discussions. January occurs twice here, and this signals that, at the beginning of most years, the agency's attention on target analysis peaked. On the contrary, we saw the lowest drops in April, June, September, and June for 2003, 2004, 2005, 2006, respectively. The double occurrence of June also indicates a reoccurring shift from target analysis discussions in that month. Although not the lowest in 2005, April was comparatively low enough in that year to merit its inclusion in this analysis. This pattern reveals that April is another month within which the agency focuses more on topics other than target analysis. Regardless of the lows and highs, there was a noticeable ascendency in target analysis-related discussions, showing that the agency gradually paid more attention.

From these trends, we could learn more about the NSA. We note that the frequent increments and decrements in the quarterly trends are likely because of the numerous topics covered in SIDtoday. The NSA is a big agency, so it is typical to expect its newsletter to contain mundane general topics of interest to its staff. In this study, we notice the agency regularly engages in discussions that run across eighteen (18) broad topics or subject areas, making it nearly impossible to, on a monthly or quarterly basis, observe consistent rise or fall in a particular topic. This consistency, however, is likely when assessing the trends yearly, as seen in Figures 5.1, 5.1.1, 7.1, and 7.1.1. A long and sustained trend is rare within a short time frame (monthly and quarterly) because each publication is likely to cover a topic area that is dissimilar from the previous.

Another thing observable from the trend is linked to Topic 12’s trend at the beginning of each year. I noticed an almost invariable upsurge for the Target Analysis topic during this time of the year, and from this trajectory and the associated newsletters, I realized that the agency often engaged in discussions about target analytic tools, evaluation of target analytic endeavors in the previous year, and strategize for the year. For instance, in the newsletter for February 10, 2004, the author indicates the NSA’s vested interest in target apparatus, future projects, and hints of how the agency evaluates those analytic apparatus by writing that:

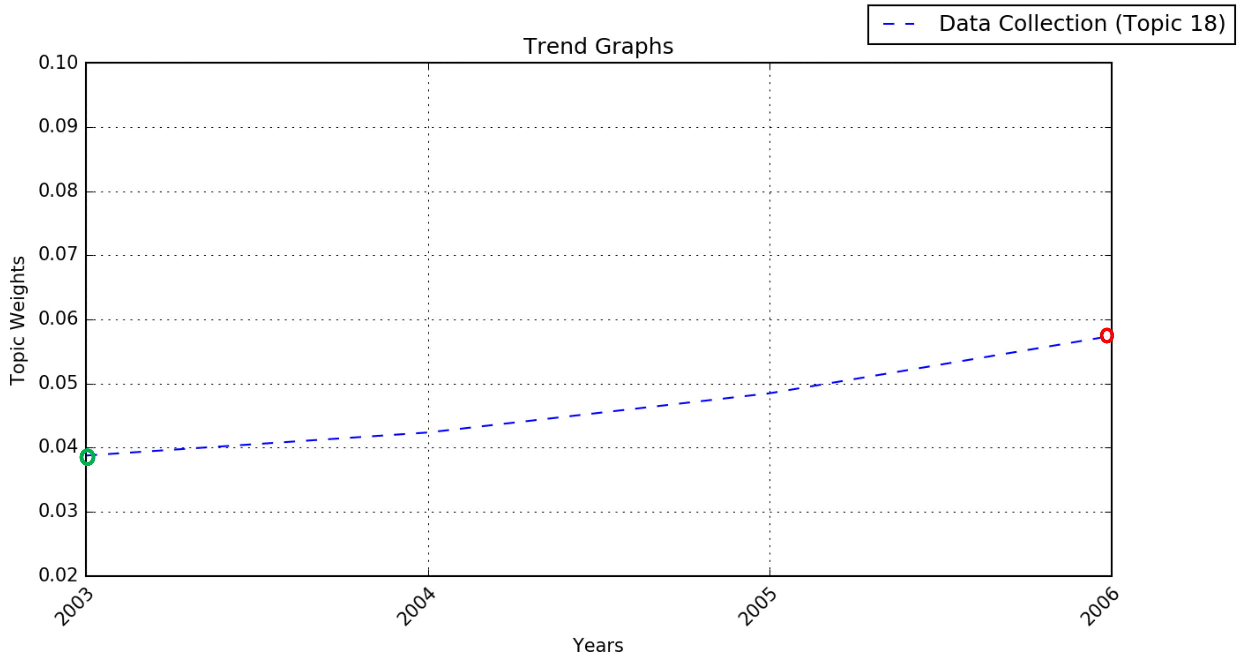
> "What technologies are our SIGINT intelligence targets using today and what will they use three years from now?" (S//SI) That's the question that the SIGINT Development Target Technology Trends Center (T3C, a.k.a. S3TSD) answers to help posture SID against strategic surprise. We rely heavily on the expertise of A&P target offices, DA technologists, private industry and academia as we track the indicators of change in the current and planned usage of technologies by our targets. The stronger the target evidence, the more certain SID is that its technology developments and investments are on track. (U//FOUO) Our primary deliverable is a Top 10 List of Target Technologies which helps drive SID's budget builds and investment portfolio. It is a living list and we are always looking for expert input.

In all, the trends for Topic 12 reveal that the NSA’s interest in analyzing targets continued to increase. It also discloses that the agency shifts its attention to other issues during certain periods in the year. Despite the undulating nature of the trend, the topic's up and down pulls reveal the agency’s serious approach to analyzing its targets, as it endeavors to engage in such discussions despite the demands of other issues (topics) within the agency.

Topic 18 (Data Collection)

Analyzing the trend of Topic 18 (Data Collection) reveals some facts about the NSA. From a yearly perspective, the trend in Figure 7.1 indicates that starting from 2003, the NSA’s attention to matters about Topic 18 (Data Collection) increased gradually, with each subsequent year showing an increase in the agency's interest. This trend shows that data gathering concerns have surged within the agency and are reflected in their SIDtoday discussions.

Figure 7. 1.
Yearly Trend Graph for Topic 18 (Data Collection).

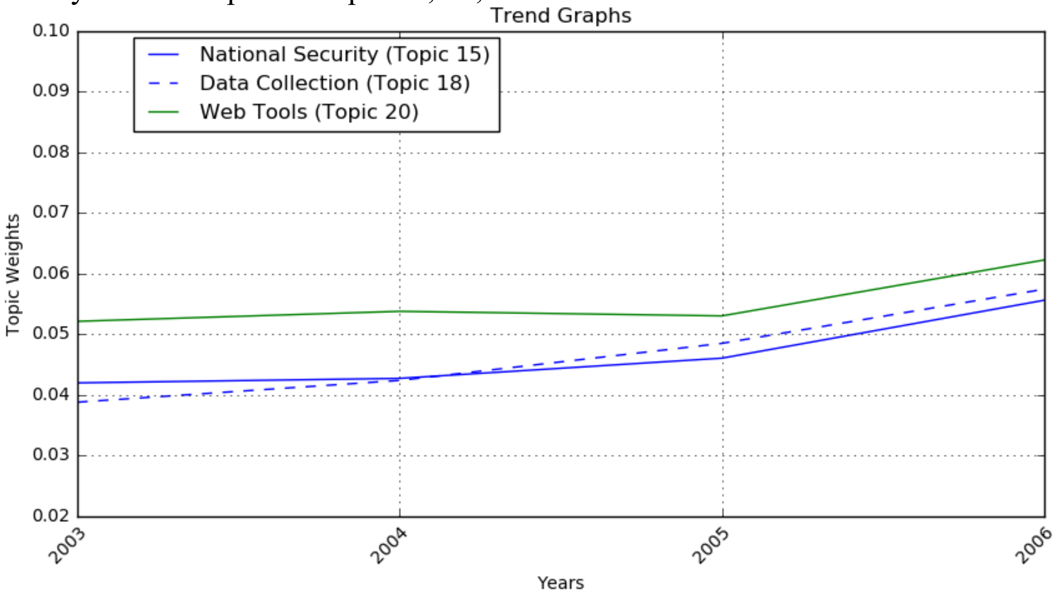


The blue dashed line indicates Topic 18's gradual rise from 0.039 (marked in green) in 2003 to 0.058 (marked in red) in 2006.

Going further into the analysis of this trend revealed that new threats and new technologies could be the main contributors to the gradual increase in the agency's attention to Data Collection topics, especially in 2006. To corroborate this hypothesis, I analyze the correlation between Topic 18 (Data Collection) and two other topics -- Topic 15 (National Security) and Topic 20 (Web Tools) -- since they both possess higher degrees of connections with new threats and new technologies. In 2006, we noticed from Figures 7.1.1 and 7.1.2 that the three (3) topics had a similar trajectory. This trajectory confirms a relationship between these topics, which asserts that the increase in NSA's new threats and technology discussions possibly impacted data collection discussions. Also, we saw those increments in Topics 15 and 20 because the agency deployed more tools in 2006 and dealt with national security concerns in the face of increased terrorism. Evidence is in SIDtoday publications for May 1, 2006, September 26, 2006, April 4, 2006, and several others.

Figure 7.1.1

Yearly Trend Graph for Topic 15, 18, and 20



In this graph we notice an obvious similarity in the trend of all the three topics, especially in 2006, suggesting a correlation between them.

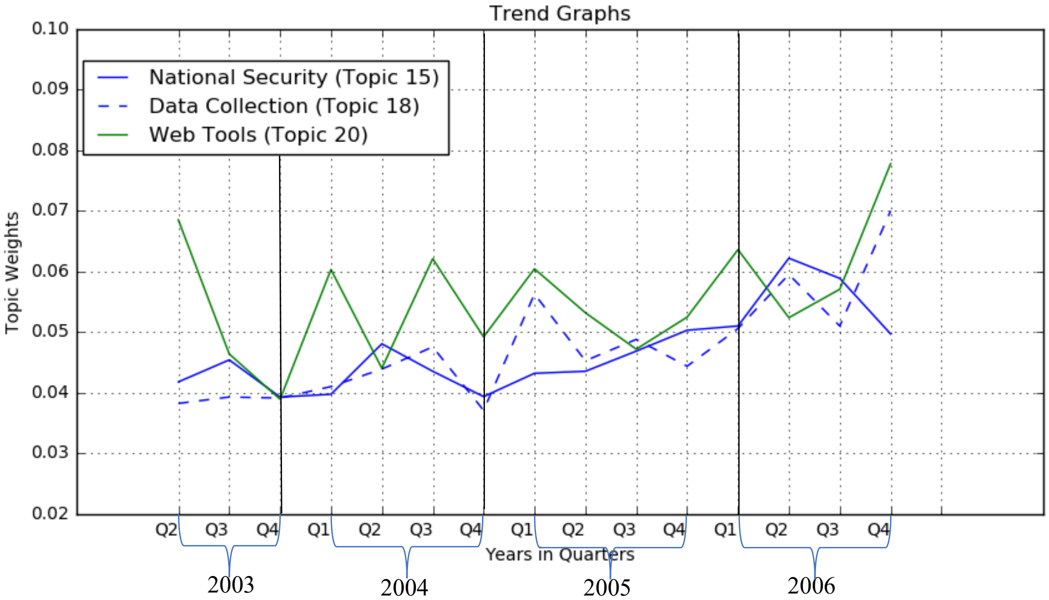
Corroborating evidence of the relationship between these three (3) topics is also apparent in other SIDtoday newsletters. For instance, I noticed discussions of new technologies (Topic 20 related) in SIDtoday's publications for May 7, 2003, July 16, 2004, July 18, 2004, September 9, 2004, December 6, 2004, February 7, 2005, November 14, 2005, January 11, 2006, January 25, 2006, October 31, 2006, November 28, 2006, December 4, 2006, December 21, 2006, and several others. I also realize that most of these newsletters are 2006 publications, explaining the rise we observe in the trend graph for 2006. From January 11, 2006's SIDtoday publication, for example, the author said:

> BLACKPEARL is a survey tool developed in the Network Analysis Center to help Characterize transmission links and the networks present on them. Currently, its most robust features provide a characterization of IP networks in a small number of packets, often around 10,000 packets. This brown bag will walk through usage of BLACKPEARL to help identify best points of access within the SIGINT system to collect on a target IP network (p. 1).

In the above, the author made us understand that the NSA had a new technology called BLACKPEARL, capable of helping them improve their data collection undertakings. Like Topic 20 (Web Tools), there is evidence indicating that Topic 15 (National Security) influenced Topic 18 in SIDtoday newsletters for January 11, 2006, January 17, 2006, July 24, 2006, December 21, 2006, and several others. In the newsletter for December 21, 2006, for instance, the author wrote that "It has been a challenging year, with the campaigns in Iraq and Afghanistan, the larger Global War on Terrorism, and a host of other issues all demanding top-notch SIGINT support (p.1)." In this context, the emphasis is on Global War on Terrorism and demanding SIGINT support because they point to the fact that new threats have impacted the high demand for SIGINT. SIGINT, almost invariably, involves the collection of data. Hence, new threats, which in many ways refers to Topic 15 (National Security), influenced the upsurge in Topic 18. Discussions and trends of this kind confirm that Topic 15 and Topic 20 affected Topic 18.

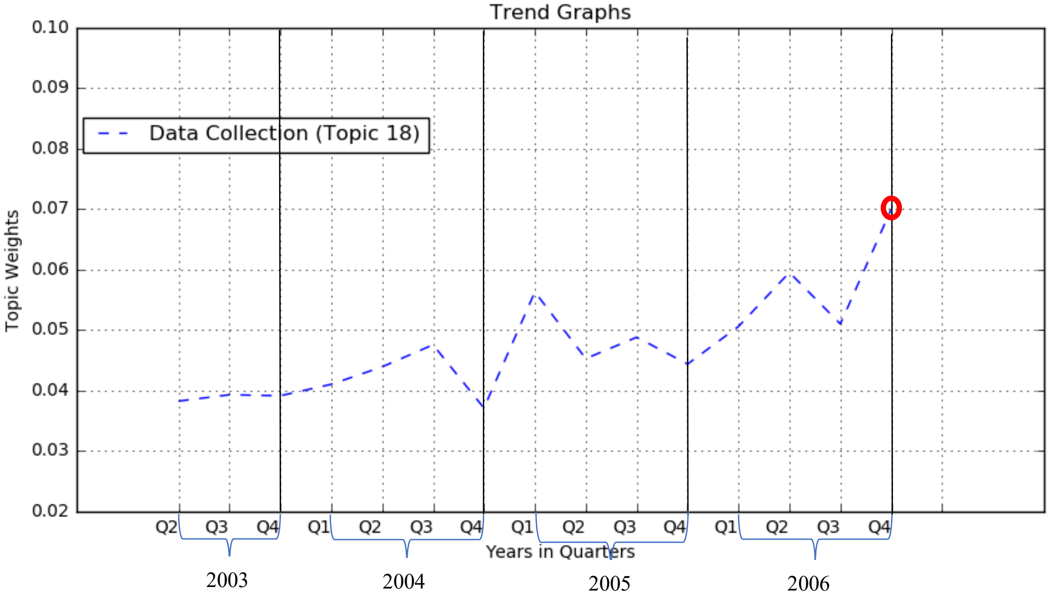
Figure 7.1.2

Quarterly Trend Graphs of Topics 15, 18, and 20.



This figure shows a quarterly view of the similarities between the three topics. The trends, although nuanced, reveal an intertwined relationship between the topics, suggesting the impact Topics 15 and 20 had on Topic 18, especially in 2006. Although concerns relating to national threats declined in mid-2006, data collection and new technology issues were still prevalent because the agency most likely aimed at preventing such occurrences in the future.

Figure 7. 2.
Quarterly Trend Graph of Topics 18



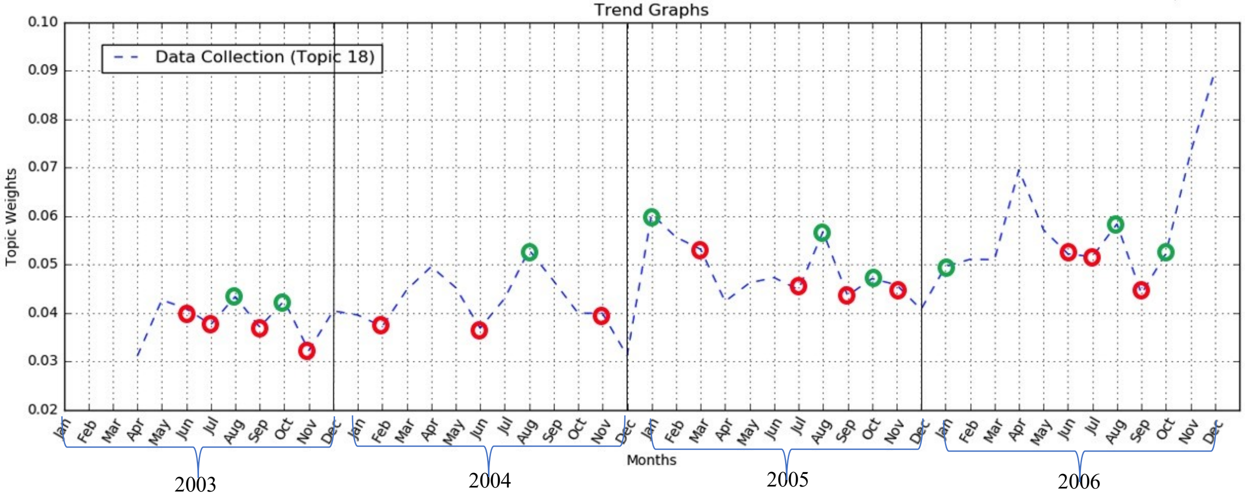
Note: Toward the end of 2006, we notice Topic 18's rise to 0.07 (marked in red), indicating that the agency's interest in the topic maintained a gradual rise throughout the years.

Moving forward, in Figure 7.2, the quarterly trend of Topic 18 tells us more about the performance of the topic within the years. Here, the trend shows that the NSA's interest and discussion of Data Collection received minimal attention for all the four quarters in 2003. The topic's attention, however, began to rise in the first quarter of 2004 and then dropped to its lowest in the fourth quarter of the same year. From there, we saw a surge in the first quarter of 2005 and a slight dip for the remaining quarters in that year.

In 2006, however, the trend continued to rise with a considerable increase from the first quarter to the second. From the third to the fourth of 2006, we saw a significant surge, ending the year with a topic weight of 0.07 (7%). Assessing the topic weights, one obvious fact we noticed about the trend was the sustained increase in the trajectory of Topic 18 (Data Collection) despite the frequent dips and surges. These trends make it conspicuous that the NSA's interest in data collection has not declined or remained steady but has continued to gain attention.

Another indication obtainable from the trend in Figure 7.2 is the minimal level of consistency in the quarters. Thus, the different quarters displayed little regularities among themselves throughout the years. For example, the first quarters (Q1) of 2003, 2004, 2005, and 2006 were disparate, revealing minuscule information about the pattern that runs across that particular quarter of the years. In the same fashion, Q2, Q3, and Q4 exhibited patterns similar to Q1's. These patterns show that the NSA pays no particular attention to what time of the year it discusses Data Collection issues, suggesting that the agency's engagement in data collection is mainly a year-wide activity.

Figure 7. 3.
Monthly Trend Graph of Topics 18.



Note: The red circles indicate months where Topic 18 decreased, while the greens indicate months when the topic increased. The black vertical lines (not the grids) section the graph by year. This trend shows that, monthly, NSA's attention to Topic 18 does not follow any specific pattern. We, however, see a gradual overall increase in Topic 18 throughout the years and towards the end of 2006.

Assessing Figure 7.3, however, reveals a trend that is similar to what we see in Figures 7.1 and 7.2. The figure shows that within the four (4) years, some months recorded frequent down-surges, and others, upsurges regarding NSA's discussion of Data Collection topics. For instance, in February, June, July, September, and November, we noticed a decline from the previous month concerning NSA’s focus on the topic. Also, in June 2004, the NSA discussed Data Collection less than it did in the month before -- May. In contrast, January, August, and October recorded more frequent increments, revealing the agency’s attention to those months increased from the previous months. Other months, nonetheless, gave a neutral indication, as they neither leaned toward declining nor inclining. Although the focus on Topic 18 fell for many months, the overall trend shows an upward trajectory, signaling that the NSA, despite the ups and downs, saw a sustained and gradual increase in interest for Data Collection discussions.

Regarding the lowest and highest points for the various months, we noticed an interesting pattern. In August, August, January, and December for 2003, 2004, 2005, 2006, respectively, Topic 18 (Data Collection) discussions recorded their peaks. Other than these trends indicating a focus on Topic 18 for the said months, it also reveals that the NSA’s tends to give the topic significant attention in some month(s) – August. On the other hand, in April, December, December, and September, for 2003, 2004, 2005, 2006, respectively, the Data Collection topic reached its lowest points. This trend also reveals discussions other than Data Collection are paramount to the NSA in some months, particularly in December. Despite these patterns in NSA’s data, we again notice an overall ascendency for Topic 18 over the four (4) years period, revealing the agency’s consistent interest in collecting data.

Similar to Topic 12’s, the frequent increments and decrements in Topic 18 are because of the high number of topics covered in SIDtoday. Due to the large number of topics covered, it is typical to see regular increments and decrements in Topic 18’s monthly and quarterly trend graphs. The yearly trend graphs, however, show a different picture as their time intervals are lengthy.

Ultimately, Topic 18’s (Data Collection) trend, throughout the four years, revealed to us an upward trajectory of issues about data gathering and surveillance within the NSA. It also gave us the insights that the agency has no preferred time frame within the year for discussing data gathering, as they engage in such discussions as to when they need to.

Limitations

Although inherently resourceful at identifying topics in discussions within a corpus, the topic modeling algorithm failed to capture the nuances or details that could otherwise be meaningful. Thus, although the topic modeling method I employed in this study enabled exploration of patterns, such as the trend of Target Analysis and Data Collection, it was unable to reliably capture the concepts, let alone understand the context of the terms.

Another limitation of this study is the data size. This study is based only on NSA's SIDtoday publications from 2003 to 2006, making the research study limited to the contents they published in the newsletters from 2003 to 2006. It is also reasonable to assume that SIDtoday does not contain all discussions employees engage in within the agency. Possibly, the agency omits outcome details from these. For instance, I expected to see in the SIDtoday's 2005 newsletters traces of Verizon's metadata phone call record sharing with the NSA, but this was not the case. This non-existence indicates that the NSA does not publish some of the more sensitive information in SIDtoday. Also, the limited number of SIDtoday newsletters available to the public has invariably affected the ability of this study to be thorough. It would have been more beneficial for this study to have a dataset that spans over a minimum of ten (10) years. Regardless, given the agency's highly secretive nature, I consider it a fortunate occurrence for such a dataset to be available in the public domain.

The number of topics (NT) requirement of the topic modeling algorithm is another factor worth considering as a limitation. As suggested by the name, NT refers to the number of themes in a corpus. It is a parameter that the researcher needs to supply the LDA algorithm before it runs, but determining this number can be a difficult task (Jacobi et al., 2016). The difficulty is that researchers who use the topic modeling algorithm hope to learn the themes in the corpus and the number of themes. Unfortunately, the algorithm requires that the researcher supplies the number of topics before it can proceed, presenting a chicken or the egg causality dilemma. As a result of difficulty, traces of doubt remain in my mind about whether my choice of twenty (20) topics was appropriate and representative of the actual topics in the corpus. Hence, the number of topics element in topic modeling is a constraint.

Aside from the challenge I encountered in determining the number of topics, there was also the issue of providing labels for the topics the model produces. To the model, a topic is a cluster of words that frequently occur together. An example of a topic is: -- "force," "military," "command," "north," "air," "korea," "army," "plan," "joint," and "service." From this list of words, a researcher(s) must determine a label or name to describe the overall theme, but this task can be error-ridden, as the process is based entirely on the researcher and his cross-validation team. For the example above, I provided "military" as its label, and so did the cross-validator. However, there is no guarantee that other researchers would not consider as more appropriate another label for this topic, ultimately giving room for inconsistencies.

Another limitation of this study is Topic Modeling's assumption that words (terms) that occur together are related, suggesting they refer to a common topic or concept. The Topic Model method I utilized in this study is dependent on terms, assessing their location or position of occurrence in a body of text and their frequency to determine whether a group of words form a topic. Through this observance of terms, topic modeling can extract topics from the text (corpus), but this approach is not always accurate. Words can occur together and yet refer to different concepts. More so, the issue of homophones and homonyms could pose challenges to topic modeling, as some terms could refer to various things or misunderstood when taken out of their original context.

In all, the limitations inherent in the data and Topic Modeling are valid but does not impede this study, as all research studies are likely to encounter unavoidable challenges. Drawbacks such as the inability to detect nuances in topics, dearth supply of data, difficulty in determining the number of topics, providing labels, and the assumption of co-occurrence are typical in unsupervised machine learning study, but these challenges present researchers opportunities to keep improving on their work.

Chapter 4: Conclusion

Surveillance is an activity of interest to the U.S., and the NSA is the agency that oversees its undertaking. To gain insight into surveillance in the U.S., this thesis aimed at identifying the themes in NSA's SIDtoday newsletters to ascertain if any of its discussions comprised data surveillance. It also focused on revealing trends that are evident in NSA's surveillance discussions. Using the Latent Dirichlet Allocation topic modeling approach, the study was able to identify eighteen (18) generally useful domains of discussions in SIDtoday. The results indicate that the NSA engages in discussions that fall under data surveillance, dissemination of information, leisure, human resource, support services, operations, intelligence community, and national security. Looking into the details of these themes reveals the agency's uniqueness as well as its mundaneness.

The analysis made it obvious that SIDtoday comprised discussions (themes) about surveillance, indicating that the agency actively monitored and talked about its targets. The LDA model revealed this surveillance culture by identifying Data Collection and Target Analysis as part of the themes in SIDtoday. And a close analysis of these themes led to the conclusion that the NSA engaged in data gathering and analysis. Using sophisticated tools and human resources, SIDtoday indicates that the NSA obtained data that they later analyze to acquire meaningful insight. This study also revealed NSA's constant innovative and collaborative efforts with other similar agencies to collect data for insights.

From the trend analysis of the NSA's Surveillance discussions (themes), this study concluded that Data Collection and Target Analysis topics (themes) had undergone some changes from 2003 to 2006. On an annual basis, the Target Analysis issues (themes) saw a gradual increase over the years, indicating the NSA's discussion of Target Analysis-related themes continued to ascend. The upward trend also signaled the NSA's gradual increase in issues about surveillance. Similarly, Data Collection also continued to increase, indicating that the agency's interest in gathering data continued to grow as the years advanced. As shown by the trend analysis, the agency expressed no particular preference for any quarter of the year concerning surveillance discussions. It appears the agency operated on a need basis.

Using a computer-assisted text analysis approach in this study was essential in turning the NSA's SIDtoday newsletters into analyzable data for this thesis. More so, a computer-assisted text analysis was relevant to this study because of its ability to effectively manipulate large volumes of textual data, such as the one in this study. This study would have been time-consuming if the traditional means of processing data were the typical approach. Noteworthy is the research possibility that a computer-assisted text analysis method like topic modeling brought to this study. Topic modeling enabled this study to identify the types of discussions that the NSA's employees engage in and the trend of those discussions. Besides, this study reached its goal of revealing prevalent themes within SIDtoday, and it achieved this goal with no disconnection between the aspired and the actual results. Also, it achieved these goals despite the technical difficulties in determining a key parameter for the LDA model. Without a doubt, this study could not have consistently and accurately extracted data about the discussions within SIDtoday without a computer-assisted approach. The process would have been excessively error-prone and time-consuming, even if several humans were responsible for this research.

Future research may analyze the various projects undertaken by the agency that have been mentioned in SIDtoday to determine the ones related to surveillance. The studies could further assess the projects based on whether the agency used them for individual or mass surveillance projects. The various names of people, locations, and objects are also opportunities for future researchers to conduct entity relationship analysis to establish connections between entities and what those links mean. Entity analysis studies involving people, locations, and other objects could be valuable for a network analysis study.

It is worth mentioning that, despite mainstream media's notions that the NSA engaged in data surveillance, this study systematically verified those claims by studying the agency's newsletters. Approaching this research with the hypothesis that the NSA discussed data surveillance, the study used natural language processing and other text analysis techniques to ascertain the agency's involvement in surveillance. But this surveillance was specifics to individuals, many of whom were terrorists. In its quest, this study has added to the systematic evaluation of the agency's activities by confirming its engagement in data surveillance, thereby contributing to the theory surrounding the NSA's surveillance culture and discourse. In a general sense, the study played a broader role in ascribing surveillance activities to any agency involved in Signal Intelligence.

Another thing is that this study revealed that the NSA is in many ways mundane and uncovered the agency's capability and active participation in monitoring people in several countries. It also shows that the agency engaged in activities that, from another perspective, could be labeled ominous but beneficial to its victim. As the most powerful country, it is expected that the United States would inevitably have enemies whose interest might be to cause America and its allies harm. And although mass surveillance is indeed invasive and wrong, the potential consequences for ignoring it may be unbearable in a world where terrorism and hostility are prevalent. Ideally, the NSA should use the proper approach to secure the required data for its

SIGINT purposes. But realistically, such an open approach may prove counterintuitive for an agency whose operation, for good reasons, is concealed from the public.

In all, the trends in the surveillance-related themes also showed no sign that the agency was decreasing its monitoring activities. The likes of Snowden, Greenwald, Poitras, and Gellman had every reason to be cautious about their online activities and public appearances. Therefore, everybody ought to be mindful of their (especially persons whose actions challenge governments) interactions online and offline. Possibly, they should take advantage of tools like Pretty Good Encryption (PGP), The Onion Router (TOR), and Virtual Private Networks (VPN) to protect themselves online. Freedom from a surveillance agency (state) means freedom to critique governments in ways that nudge them to give their best to those they serve.

List of References

About the Intercept. (2019). Retrieved September 20, 2007, from https://theintercept.com/about/

Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. https://doi.org/10.1007/978-3-319-14142-8

Aid, M. M. (2001). The National Security Agency and the Cold War. *Intelligence & National Security*, 16(1), 27-66. https://doi.org/10.1080/02684520412331306200a

Aid, M. M. (2009). *The secret sentry: the untold history of the National Security Agency*. 1st U.S. ed. New York: Bloomsbury Press.

Aid, M. M., & Wiebes, C. (Eds.). (2001). Secrets of signals intelligence during the Cold War and beyond. *Psychology Press*.

Akef, I., Arango, J. S. M., & Xu, X. (2016). Mallet vs GenSim: Topic modeling for 20 news groups report. Retrieved from https://www.researchgate.net/profile/Islam-Ebeid/publication/331972126_Mallet_vs_GenSim_Topic_Modeling_Evaluation_Report/links/5c96e229a6fdccd46036707e/Mallet-vs-GenSim-Topic-Modeling-Evaluation-Report.pdf

Albrecht, L. (2005). *Textual Analysis and the Production of Text*. Samfundslitteratur.

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society, 12(2),* 185-196.

Ball, K., Di Domenico, M., & Nunan, D. (2016). Big Data Surveillance and the Body-subject. *Body & Society*, 22(2), 58–81. https://doi.org/10.1177/1357034X15624973

Bamford J. (1982). *The Puzzle Palace: A Report on America's Most Secret Agency*. Boston, MA: Houghton Mifflin.

Bamford, J (2012). The NSA Is Building the Country's Biggest Spy Center (Watch What You Say). Retrieved from https://www.wired.com/2012/03/ff-nsadatacenter/

Bamford, J. (1983). *The Puzzle Palace: Inside the National Security Agency*, America's Most Secret Intelligence Organization. Granite Hill Publishers.

Bamford, J. (2001). *Body of secrets: Anatomy of the Ultra-Secret National Security Agency: from the Cold War through the dawn of a new century*. New York: Doubleday.

Bamford, J. (2002). *Body of secrets: Anatomy of the Ultra-Secret National Security Agency*. 1st Anchor Books ed. New York: Anchor Books.

Bamford, J. (2007). *Body of secrets: Anatomy of the ultra-secret National Security Agency*. Anchor.

Bamford, J. (2018). *The Puzzle Palace: a report on NSA, America's most secret agency*. Houghton Mifflin Harcourt.

Banks, W. C. (2006). The Death of FISA. Minn. L. Rev., 91, 1209. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/mnlr91&div=39&g_sent=1&casa_token=wDttI87-lF8AAAAA:wAlxjdsKGJex1pbnlOXm1paUS30CBsmn2pQ0ehwTWl-Q5gzNWmbsw2r34sYHr2WWcMzfaklftg&collection=journals

Barde, B. V., & Bainwad, A. M. (2017, June). An overview of topic modeling methods and tools. In 2017 *International Conference on Intelligent Computing and Control Systems* (ICICCS) (pp. 745-750). IEEE.

Bartlett, J. (2015). *The dark net: inside the digital underworld*. Brooklyn: Melville House.

Benson, R. L. (1997). *A History of US Communications Intelligence during World War II: Policy and Administration.* Center for Cryptologic History, National Security Agency.

Benson, R. L., & Warner, M. (1996). *Venona: Soviet Espionage and the American Response 1939-1957.* National Security Agency.

Bernard, R. L. (2009). Electronic Intelligence (Elint) at NSA. Retrieved from https://repository.library.georgetown.edu/bitstream/handle/10822/1053203/elint.pdf?sequence=1

Bhatia, P. (2019). *Data Preprocessing. In Data Mining and Data Warehousing: Principles and Practical Techniques (pp. 55-64).* Cambridge: Cambridge University Press. doi:10.1017/9781108635592.005

Blake, A. (2013, June 10). Edward Snowden apparently a Ron Paul supporter. *The Washington Post.* Retrieved from https://washingtonpost.com/news/post-politics/wp/2013/06/10/edward-snowden-apparently-a-ron-paul-supporter/?noredirect=on&utm_term=.cda0fe6642c7

Blei, D. M. (2012). Topic modeling and digital humanities. *Journal of Digital Humanities*. Retrieved from http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/

Breckinridge, S. D. (1993). *The CIA and the Cold War: a memoir*. Praeger Pub Text.

Burns, T. L. (2005). The Origins of the National Security Agency. *United States cryptologic history*, 1. Retrieved from https://tucops.info/tucops3/etc/spies/originsa.pdf

Burrows, W. E. (1996). Imaging Space Reconnaissance Operations during the Cold War: Cause, Effect, and Legacy. U-2 Flights and the Cold War in the High North, 84.

CAIR-NY. (2013, May 16). Glenn Greenwald Speaks at CAIR-NY Annual Banquet. *CAIR-NY*. Retrieved from https://www.cair-ny.org/news/2013/5/16/glenn-greenwald-speaks-at-cair-ny-annual-banquet

Cauley, L. (2006). NSA has massive database of Americans' phone calls. *USA today*, *11*(06).

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems, 22, 288-296.*

Clarke, R. W. (2003, March). Dataveillance-15 years on. *In Privacy Issues Forum (Vol. 28).* Retrieved from http://www.rogerclarke.com/DV/DVNZ03.html

Culture Shock: NSA from the Perspective of Summer Interns. (2011, July 27). Retrieved from https://assets.documentcloud.org/documents/2830624/2011-7-27-Culture-Shock-NSA-From-the-Perspective.pdf

Debortoli, S., Müller, O., Junglas, I., & vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems, 39(1), 7.*

Durham, M. (2000). The Christian Right, the Far Right and the Boundaries of American Conservatism. Manchester, UK: *Manchester University Press*.

Dynamic Methods of Interaction with New and Existing Customers. (2003). Retrieved from https://theintercept.com/snowden-sidtoday/2829992-dynamic-methods-of-interaction-with-new-and/

EMC Education Services (Ed.). (2015). Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. *Wiley.* https://doi.org/10.1002/9781119183686

Epstein, E. J. (2017). *How America lost its secrets: Edward Snowden, the man and the theft (1st ed.).* New York: Alfred A. Knopf.

Espionage, E. (2013). the NSA’s Secret Spy Hub in Berlin. Spiegel Online: International. Retrieved from https://www.spiegel.de/international/germany/cover-story-how-nsa-spied-on-merkel-cell-phone-from-berlin-embassy-a-930205.html

Evans, M. S. (2014). A computational approach to qualitative analysis in large textual datasets. *PloS one*, *9*(2), e87908.

Fantz, A., Black, P., & Martinez, M. (August 2013). Snowden out of airport, still in Moscow. *CNN.* Retrieved from edition.cnn.com/2013/08/01/us/nsa-snowden/index.html

Ferran, L. (2013, June 12). Snowden’s CIA Drunk Driving Claim Questioned - ABC News. *ABC News*. Retrieved from https://abcnews.go.com/blogs/headlines/2013/06/switzerland-questions-u-s-over-cia-drunk-driving-gambit/

Finn, P., Miller, G., & Nakashima, E. (2013). Probe aims to find how NSA leaker got access. Pittsburgh Post-Gazette. Retrieved from http://post-gazette.com/news/nation/2013/06/11/Probe-aims-to-find-how-NSA-leaker-got-access/stories/201306110140

Fitsanakis, J. (2007). National Security Agency: The historiography of concealment. In The History of Information Security (pp. 523-563). *Elsevier Science BV*. doi://doi.org/10.1016/B978-044451608-4/50019-5

Gellman, B & Poitras, L (2013, June 7). U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program. *The Washington Post*, Retrieved from https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html?noredirect=on&utm_term=.d5b728956fbe

Gellman, B. (2020). *Dark Mirror: Edward Snowden and the American Surveillance State*. Penguin Press.

Gellman, B., & Poitras, L. (2013). US, British intelligence mining data from nine US Internet companies in broad secret program. *The Washington Post*, 6(07), 2013.

Gilbert, J. L., & Finnegan, J. P. (Eds.). (1993). US Army signals intelligence in World War II: a documentary history (Vol. 70, No. 43). *US Government Printing Office*.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., ... & Smith, N. A. (2010). *Part-of-speech tagging for twitter: Annotation, features, and experiments.* Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

Glatzer, J., & Poitras, L. (2006). My country my country. [New York, N.Y.]: Zeitgeist Films.

Gordon, J. (2020, December). 'The greatest gift of all': Whistleblower Edward Snowden and his wife Lindsay Mills announce birth of baby boy with festive photographs. *Daily Mail.* Retrieved from www.dailymail.co.uk/news/article-9088131/Whistleblower-Edward-Snowden-wife-Lindsay-Mills-reveal-pics-new-baby.html

Goulden, J. C. (1969). Truth is the first casualty: The Gulf of Tonkin affair: illusion and reality. Rand McNally.

Greenberg, A. (2014, September 13). These are the Emails Snowden Sent to First Introduce His Epic NSA Leaks. *Wired*. Retrieved from https://www.wired.com/2014/10/snowdens-first-emails-to-poitras/

Greene, D., O'Callaghan, D., & Cunningham, P. (2014, September). How many topics? stability analysis for topic models. In Joint European conference on machine learning and knowledge discovery in databases (pp. 498-513). *Springer, Berlin, Heidelberg*.

Greenwald, G & MacAskill, E (2013, June 7). NSA Prism program taps into user data of Apple, Google and others. *The Guardian*, Retrieved from https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data

Greenwald, G. (2014). No Place to Hide. Edward Snowden, The NSA and The Surveillance State. *Penguin Books*.

Greenwald, G., & Ackerman, S. (2013). NSA collected US email records in bulk for more than two years under Obama. *The Guardian, 27.*

Greenwald, G., & MacAskill, E. (2013). NSA Prism program taps into user data of Apple, Google and others. *The Guardian, 7(6), 1-43.*

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism and Mass Communication Quarterly*. https://doi.org/10.1177/1077699016639231

Gurnow, M. (2014). *The Edward Snowden affair: exposing the politics and media behind the NSA scandal*. Indianapolis: Blue River Press.

Hildebrandt, A (2013). Why the U.S. hasn't nabbed Edward Snowden yet. *CBC News*. Retrieved from www.cbc.ca/news/canada/why-the-u-s-hasn-t-nabbed-edward-snowden-yet-1.1310160

History of the Signal Security Agency (1947). The Japanese Army Problems – Cryptanalysis. (ASA 1947). Holding Area, Accession No. 2465, CBIB 14, NSA (S).

Horgan, P. S. (1991). *Signals Intelligence Support to US Military Commanders: Past and Present*. Army War Coll Carlisle Barracks Pa.

Howe, G. F. (1974). The Early History of NSA. *Cryptologic Spectrum,* 4(2), 11-17.

Ilyushina, M (2020, October). Edward Snowden gets permanent residency in Russia. *CNN*. Retrieved from www.cnn.com/2020/10/22/europe/edward-snowden-russia-residency-intl/index.html

Inman, B. R. (1979). The NSA perspective on telecommunications protection in the nongovernmental sector. *Cryptologia*, 3(3), 129-135.Intelligence & National Security, 16:1, 27-66, DOI: 10.1080/02684520412331306200a

Intelligence Community Classification Guidance Findings and Recommendations Report. (2008). Retrieved from https://fas.org/sgp/othergov/intel/class.pdf

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital journalism, 4*(1), 89-106.

Jockers, M. L., & Thalken, R. (2014). *Text analysis with R for students of literature*. New York: Springer.

Johnson, T. R. (1995). American cryptology during the Cold War, 1945-1989. Center for *Cryptologic History*, National Security Agency.

Johnson, T. R. (2015). Pearl Harbor. St. Catharines, Ontario: Crabtree Publishing Company.

Johnson, T. R., & Hatch, D. A. (1998). NSA and the Cuban Missile Crisis. *National Security Agency*, May.

Kilian, C. (2014, May). Why Edward Snowden Chose Glenn Greenwald. *The Tyee*. Retrieved from https://thetyee.ca/Books/2014/05/30/No-Place-to-Hide/

King, R. (2014). Ex-NSA Chief Details Snowden's Hiring at Agency, Booz Allen. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/exnsa-chief-details-snowden8217s-hiring-at-agency-booz-allen-1391569429

Klein, M., & Bamford, J. (2009). *Wiring Up the Big Brother Machine--and Fighting it*. BookSurge.

Koenig. (1993). Koenig Solutions. Retrieved May 12, 2019, from https://www.koenig-solutions.com/about-koenig

Lee, J. H., Kim, Y. G., & Yu, S. H. (2001, January). Stage model for knowledge management. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (pp. 10-pp). IEEE.

Lee, M. (2020). Pytesseract. Retrieved June 2, 2020, from https://pypi.org/project/pytesseract/

Leydesdorff, L., & Nerghes, A. (2017). Co-word maps and topic modeling: A comparison using small and medium-sized corpora (N< 1,000). *Journal of the Association for Information Science and Technology, 68*(4), 1024-1035.

Li, B., Erdin, E., Güneş, M. H., Bebis, G., & Shipley, T. (2011, April). An analysis of anonymizer technology usage. In International Workshop on Traffic Monitoring and Analysis (pp. 108-121). Springer, Berlin, Heidelberg.

Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An Overview of Topic Modeling and its Current Applications in Bioinformatics. *SpringerPlus*, 5(1), 1608.

Loeb, V. (2001). Test of Strength. Washington Post Magazine, W08.

Loiko, S. L. (2015). Snowden stopping in Moscow en route to Cuba, Russian says. *Los Angeles Times*. Retrieved from www.latimes.com/world/la-xpm-2013-jun-23-la-fg-wn-snowden-expected-in-moscow-20130623-story.html

Maass, P. (2013, August). How Laura Poitras Helped Snowden Spill His Secrets. *The New York Times Magazine*. Retrieved from https://www.nytimes.com/2013/08/18/magazine/laura-poitras-snowden.html

Maass, P. (2016, March 16). What it's Like to Read the Nsa's Newspaper for Spies. *The Intercept*. Retrieved from https://theintercept.com/2016/05/16/what-its-like-to-read-the-nsas-newspaper-for-spies/

Marbella, J., Bengali, S., & Cloud, D. S. (2013, June 10). Details about Edward Snowden's life in Maryland emerge. *The Baltiore Sun*. Retrieved from http://articles.baltimoresun.com/2013-06-10/news/bs-md-snowden-profile-20130610_1_anne-arundel-county-arundel-high-the-guardian

McAvoy, N. (2010). *Coded Messages: How the CIA and NSA Hoodwink Congress and the People*. Algora Publishing.

McCallum, A. K (2002). MALLET: A Cachine Learning for Language Toolkit. Amherst, MA, USA: University of Massachusetts. *http://mallet. cs. umass. edu*

McInnis, T. N. (2009). *The Evolution of the Fourth Amendment*. Lexington Books.

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993). A semantic concordance. In Human Language Technology: *Proceedings of a Workshop Held at Plainsboro*, New Jersey, March 21-24, 1993.

Milone, M. (2002). Hacktivism: Securing the National Infrastructure. The Business Lawyer, 58(1), 383–413. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/busl58&div=21&g_sent=1&casa_token=wVv1DG3oTDMAAAAA:-w97xfybMHLIqZ0_TdAQ0oN_CC4USVdrZc51yCRVjBxAhHObBB8cgbCDBeNzTN9fKVcNCpPZ0w&collection=journals

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 262-272).

Minnich, A., Abu-El-Rub, N., Gokhale, M., Minnich, R., & Mueen, A. (2016, August). ClearView: Data cleaning for online review mining. *In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 555-558). IEEE.

Nash, T. (2013, July 25). Welcome to Anchory. *MuckRock*. Retrieved from https://www.muckrock.com/news/archives/2013/jul/25/anchory-documents-offer-glimpse-90s-era-nsa-intell/

National-Cryptologic-School. (2006, September 25). Write Right: Breaking an Old Reporter's Heart. *SIDtoday*.

Nield, T. (2019). An Introduction to Regular Expressions. O'Reilly Media, Inc. *https://www.oreilly.com/content/an-introduction-to-regular-expressions/*

NSA/CSS. (n.d.). Mission &amp; Vision. Retrieved November 10, 2018, from https://www.nsa.gov/about/mission-values/

Number 9… Number 9... (2012, March 30). SIDtoday. Retrieved from https://www.documentcloud.org/documents/2830625-2012-03-30-SIDToday-Number-9-Number-9.html

Obar, J., & Clement, A. (2013). Internet Surveillance and Boomerang Routing: A Call for Canadian Network Sovereignty. *Proceedings of the Technology & Emerging Media Track - Annual Conference of the Canadian Communication Association (Victoria, June 5-7, 2012)*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2311792

Odd Jobs Before NSA. (2005, September 2). Retrieved from https://www.documentcloud.org/documents/2830622-2005-09-02-SIDToday-Odd-Jobs-Before-NSA-Part-2.html

Poitras, L., & Greenwald, G. (2013). NSA whistleblower Edward Snowden: “I don’t want to live in a society that does these sort of things” [video]. *The Guardian*. Retrieved from https://www.theguardian.com/world/video/2013/jun/09/nsa-whistleblower-edward-snowden-interview-video

Practical Jokes and April Fools. (2003). *SIDtoday*. Retrieved from https://theintercept.com/snowden-sidtoday/2829959-practical-jokes-and-april-fools/

Ratnaparkhi, A. (1996). A maximum entropy model for part-of-speech tagging. *In Conference on empirical methods in natural language processing.*

Rich, N. (2010, December). Meet the Most Dangerous Man in Cyberspace: The American Face of Wikileaks. *Rolling Stone Magazine.* Retrieved from http://www.rollingstone.com/culture/news/17389/238944

Risen, J. (2013, October 17). Snowden says he took no secret files to Russia. *New York Times News.* Retrieved from https://lasvegassun.com/news/2013/oct/17/snowden-says-he-took-no-secret-files-russia/

Rish, I., Cecchi, G. A., Lozano, A., & Niculescu-Mizil, A. (Eds.). (2014). *Practical applications of sparse modeling.* MIT Press.

Robertson, A. (2014, April 17). United States of Secrets’ promises “definitive history” of domestic surveillance. *The Verge.*

Rockwell, G., & Sinclair, S. (2016). *Hermeneutica: Computer-assisted interpretation in the humanities.* MIT Press.

Russell, M. A. (2013). *Mining the Social WebMining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More.* O'Reilly Media, Inc. (2nd ed.). O'Reilly Media, Inc.

Sarkar, D. (2019). *Text analytics with Python: a practitioner's guide to natural language processing.* Apress.

Schmitt, E. (2013, October 11). C.I.A. Disputes Early Suspicions on Snowden. *New York Times*.

Schofield, A., & Mimno, D. (2016). Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics.* https://doi.org/10.1162/tacl_a_00099

Shinyama, Y. (2019). PDFMiner. Retrieved June 1, 2019, from https://pypi.org/project/pdfminer/

Sidtoday. (2012, March 30). Number 9 Number 9. Retrieved from https://www.documentcloud.org/documents/2830625-2012-03-30-SIDToday-Number-9-Number-9.html

Snowden Archive. (n.d.). Retrieved January 20, 2018, from https://theintercept.com/snowden-sidtoday/

Snowden, E. (2019). *Permanent Record*. Metropolitan Books

Stein, B., & Zu Eissen, S. M. (2004, December). Topic identification: Framework and application. In *Proceedings of the International Conference on Knowledge Management* (Vol. 399, pp. 522-531).

Steury, D. P. (1996). *Intentions and Capabilities: Estimates on Soviet Strategic Forces, 1950-1983.* History Staff, Center for the Study of Intelligence, Central Intelligence Agency.

Tang, J., Zhang, M., & Mei, Q. (2013). One theme in all views: Modeling consensus topics in multiple contexts. *In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2487575.2487682

Toman, M., Tesar, R., & Jezek, K. (2006). Influence of word normalization on text classification. *Proceedings of InSciT*, *4*, 354-358.

Tor FAQ (n.d). n/a. Retrieved December 20, 2018, from https://2019.www.torproject.org/docs/faq#WhyCalledTor

Törnberg, A., & Törnberg, P. (2016). Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society*, *27*(4), 401-422.

Toxen, B. (2014). The NSA and Snowden: Securing the All-Seeing Eye. Queue, 12(3), 40. https://doi.org/10.1145/2602649.2612261

Vladeck, S. I. (2015). The FISA Court and Article III. Wash. & Lee L. Rev., 72, 1161. *Law Journal Library*. Chicago. Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/waslee72&div=26&g_sent=1&casa_token=uK_jtxlVB_wAAAAA:1k4PPX3OZaraASQRApDUIJJ-50Dsv3oZHtRikHI-xvPdFVJaOIf8DaJFbh-6uHErVdznGUPbbw&collection=journals

Watson, S. (2014, August 13). Snowden: There’s A “Holy Sh_T” Smoking Gun Revelation Coming. I*nfowars*. Retrieved from https://www.infowars.com/snowden-theres-a-holy-sh_t-smoking-gun-revelation-coming/

Welcome to SIDtoday. (2003). *SIDtoday*. Retrieved from https://theintercept.com/snowden-sidtoday/2829957-welcome-to-sidtoday/

Wiley, R. (2006). ELINT: *The interception and analysis of radar signals*. Artech House. Boston.

Xu, J. (2018). Topic Modeling with LSA, PLSA, LDA & lda2Vec. *NanoNets*, on Medium, May, 25.

Zaleski, A. (2013). Surveillance is a necessary ‘hornet’s nest’: NSA Director Keith Alexander. Retrieved from https://technical.ly/baltimore/2013/11/01/surveillance-necessary-hornets-nest-nsa-director-keith-alexander/

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015, December). A heuristic approach to determine an appropriate number of topics in topic modeling. *In BMC bioinformatics* (Vol. 16, No. 13, p. S8). BioMed Central.

Appendix A

The table below contains all the twenty (20) topics the topic modeling algorithm generated from NSA's SIDtoday newsletters from 2003 to 2006.

Topics with their Labels, Representation, and Term

Dominant Topic	Topic Label	% Total Docs	Topic Desc
Topic 1	Field Work Leisure	3.77	world, life, remember, large, live, drive, city, building, day, work
Topic 2	Session or Conference	8.35	conference, session, sinio, attend, discussion, present, presentation, registration, council, development
Topic 3	Operations	4.01	operation, mission, support, center, provide, target, element, area, build, personnel
Topic 4	Military	2.92	force, military, command, north, air, korea, army, plan, joint, service
Topic 5	History	3.16	view, history, story, president, event, insider, ambassador, send, force, turn
Topic 6	Customer Service	8.11	customer, process, plan, system, requirement, management, enterprise, mission, strategy, focus
Topic 7	Information Sharing with Partners	5.47	information, partner, share, policy, security, party, foreign, office, relationship, access
Topic 8	Leadership	3.11	make, leader, good, people, technical, leadership, idea, decision, problem, important
Topic 9	Intelligence Community	4.43	intelligence, information, analysis, community, national, analyst, issue, analytic, cia, include
Topic 10	Deployment Support	4.95	iraq, team, day, baghdad, part, iraqi, deploy, support, time, hour
Topic 11	NA	5.37	work, day, year, people, time, make, give, good, great, job
Topic 12	Target Analysis	8.25	target, network, analyst, analysis, tool, technology, datum, development, number, capability
Topic 13	Position	2.88	work, senior, time, office, agency, manager, position, officer, change, line
Topic 14	Writing	5.7	report, article, write, read, comment, question, reporting, word, product, find
Topic 15	National Security	4.24	terrorist, state, threat, security, government, attack, terrorism, war, group, east
Topic 16	Awards	3.58	year, award, member, organization, community, learn, annual, international, recognize, participate
Topic 17	Analysists' Skillset	4.29	language, program, training, analyst, skill, level, linguist, cryptologic, high, knowledge
Topic 18	Data Collection	5.04	collection, site, signal, system, communication, team, effort, survey, gchq, result
Topic 19	NA	5.04	director, deputy, staff, chief, meeting, issue, visit, directorate, quirk, provide
Topic 20	Web tools	7.31	information, web, page, user, document, cpe, contact, account, database, tool