ENHANCING OPEN-SOURCE INTELLIGENCE (OSINT) THROUGH
WEB SERVICES

A Thesis
Presented to the
Faculty of
San Diego State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
in
Homeland Security

by
Sergio Gerardo Valdez
Summer 2011
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Sergio Gerardo Valdez:

Enhancing Open-Source Intelligence (OSINT) Through Web Services

Eric G. Frost, Chair
Homeland Security Program

Cezar M. Ornatowski
Department of Rhetoric and Writing Studies

Shawn T. Flanigan
School of Public Affairs

4/26/2011
Approval Date
DEDICATION

I dedicate this thesis to my family for their love and support. Thank you.
ABSTRACT OF THE THESIS

Enhancing Open-Source Intelligence (OSINT) Through Web Services
by
Sergio Gerardo Valdez
Master of Science in Homeland Security
San Diego State University, 2011

This thesis will illustrate how “web services” (Facebook, Twitter, YouTube, Google, Bing, Yahoo, Alerts, RSS Feeds, Web Slices, and Google Reader) can and will enhance Open-Source Intelligence (OSINT) collection and utilization by the Department of Homeland Security (DHS), Intelligence Communities (IC) and its Domestic Partners (DP) that have the sole responsibility of protecting the nation. Using these web services that are already available can lead to positive outcomes for: counter-cartel operations in the US and Mexico, Middle East unrest, Humanitarian Assistance Disaster Relief (HADR), and eDiplomacy (State Department). This thesis will also demonstrate the use of these web services and how DHS can process millions to billions of pieces of information per day and potentially build data infrastructures and workflows to assist in the broad range of DHS responsibilities.

These web services will ease the collection and utilization of OSINT if managed properly, maximize relevant and critical information that can be translated into rapid intelligence, and minimize cost measures (since these web services are already built and currently in use in other forms). If these web services are not taken into consideration as an enhancement tool for OSINT, then DHS, IC, and its DP will be missing important rhetoric information (online), the proliferation of terrorism activity inside and outside the US, threats to critical infrastructure, and will be denying citizens the protection and transparency they deserve from DHS. Now, how is this actually being done? How much further behind is the US because the materials for OSINT have exploded (online)? What tools as web services are really there for the IC, for Law Enforcement or groups such as DHS? These questions among others, in conjunction with historical context of OSINT, will be answered in this thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 OSINT IS NOT NEW IN THE INTELLIGENCE COMMUNITIES (IC)</td>
<td>9</td>
</tr>
<tr>
<td>3 OPEN SOURCE INTELLIGENCE</td>
<td>15</td>
</tr>
<tr>
<td>4 HARNESSING THE POWER OF THE INTERNET AND ITS WEB SERVICES</td>
<td>26</td>
</tr>
<tr>
<td>Web Technologies: Alerts, RSS Feeds, Web Slices, and Readers</td>
<td>30</td>
</tr>
<tr>
<td>Web Services - Analysis</td>
<td>37</td>
</tr>
<tr>
<td>5 CONCLUSION</td>
<td>39</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>44</td>
</tr>
<tr>
<td>APPENDIX</td>
<td></td>
</tr>
<tr>
<td>SOCIAL MEDIA AND UNREST IN #EGYPT</td>
<td>47</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. OSINT – Benefits vs. Liabilities.................................................................19
Table 2. ACSI Internet Portals and Search Engines, News and Information, and Social Media Websites.................................................................29
Table 3. Internet Web Services.............................................................................35
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lyrics to “El Corrido de Tony Tormenta.”</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>The Foreign Broadcast Monitoring Service (FBMS) listening to broadcast information from countries.</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>World map of social networks.</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>Internet site categories – DHS can analyze for OSINT</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>Google Alerts interface.</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>Yahoo! Alerts interface.</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>Bing Alerts interface.</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>Add a Web Slice to your toolbar.</td>
<td>33</td>
</tr>
<tr>
<td>9</td>
<td>Google Reader (user interface) with subscriptions (know as RSS Feeds).</td>
<td>34</td>
</tr>
<tr>
<td>10</td>
<td>Google Alert.</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>Yahoo Alert.</td>
<td>36</td>
</tr>
<tr>
<td>12</td>
<td>Bing Alert.</td>
<td>37</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to thank Dr. Eric Frost, Dr. Cesar Ornatowski, and Dr. Shawn Flanigan for their guidance and support. Thank you.
CHAPTER 1

INTRODUCTION

How can we enhance Open-Source Intelligence (OSINT) collection and utilization, and what might be emerging opportunities for the Department of Homeland Security (DHS) and its many intelligence counterparts to more effectively fulfill their lawful responsibilities to the nation? Since the widespread deployment of the Internet in the 1990s, the Internet has distinguished itself as one of the most remarkable developmental apparatuses on the planet, functionally changing business, society, and most activities with new types of technologies, forms of connectivity, and millions of applications. Just as the very identity of what “Internet” means has dramatically changed over time, its significance for contributing to OSINT has dramatically changed. Two aspects of this are the extraordinary physical connectivity of billions of machines via fiber, wired, and wireless networks and the even more extraordinary ways in which these machines are used by billions of people and their software applications to convey text, data, imagery, video, and many other types of content.

The magnitude of what OSINT stands for and what it can provide the DHS intelligence cycle is enormous. Traditionally OSINT was commonly understood to be “that it is not classified and is available to the general public.”1 That really meant that OSINT materials were derived from newspapers, public documents, book journals, periodicals, radio, television, commercial databases, and gray literature like think-tank studies, symposium reports, academic research, corporate brochures and reports, and conference proceedings that are free of proprietary or copyright restrictions.2

As the evolution of the Internet has developed to include billions of websites and other electronic communications, so OSINT has changed its identity to include electronic materials that also fit into the same general categories of “not classified and available to the

---

1 P. F. Wallner, “Open Sources and the Intelligence Community: Myths and Realities,” American Intelligence Journal 14, nos. 2-3 (Spring-Summer 1993): 19.
2 Ibid.
general public.”3 Obviously this has taken place in some ways with the massive use of the Internet, Google searches, and now social networking sites such as Twitter and Facebook. These tools clearly make OSINT more productive, effective, usable, and pliable over time. These tools have fundamentally changed how OSINT is done, so that the current techniques for gathering and utilizing OSINT by groups such as the Office of the Director of National Intelligence where functionally the only way possible to accomplish many of the goals of the Intelligence Community (IC) is through the widespread use of OSINT.4

However, despite the huge potential of how these electronic media and applications could enhance OSINT, the sheer immensity of this new data stream and the dynamic, ever-changing character of it are staggering challenges. As a simple example, a website for a group of interest to DHS is that of Aryan Nations.5 As a continually changing web page, gleaning appropriate data from it is impossible without enormous work in following the page and its almost ever-changing content. Tools and techniques do exist, but this page has changed dramatically in the last decade, so that finding information, trends of information, and what was known or changed at what time means that OSINT on what should be a fairly straightforward interest can actually be dramatically more difficult with the current fluid nature of the Internet content. Capturing this fluidity can be done in a number of easy and difficult ways, but it makes OSINT collection a much more daunting task than with traditional media like newspapers or journals.

One of the obvious goals is to not re-invent or rediscover the wheel by dozens of people doing exactly the same thing, such as seeking to find the same information rather than utilizing the efforts of other members of the IC and DHS. Advancing the capability, or standing on the shoulders of previous workers is obviously dramatically attractive if it can be done. Efforts such as Wikipedia that allow multiple people to contribute to the end product of open-source data has obviously been duplicated within the IC with Intellipedia,6 where

3 Ibid.
5 For more information, see Aryan Nation, http://www.aryan-nations.org
6 For more information on Intellipedia, see D. C. Andrus, “The Wiki and the Blog: Toward a Complex
members of the community can share knowledge and build a data base of information rather than always starting over from scratch.

This is functionally using web services, or the distributed servers that collectively empower Wikipedia or Intellipedia to work. Instead of gathering all the data, simple links or meta-data to the real data can be included, such as a link to a website, document, YouTube video, or other content. The web services are the sharing, simply gathering the information so the wheel doesn’t have to be invented again, especially after specific work to solve a problem where the solution can be maintained and provided to others in appropriate ways as within Intellipedia according to security classifications.

Functionally then, OSINT collected and maintained over time can yield far more information than static information since trending, changes through time, visual analytics, and dozens of other attributes can be developed and used. This is like making animations of still images, yielding a movie. Thousands of still images versus a movie yield a dramatically different communication product, especially then linking music and sound to the data and information. Viewing the Internet as a continuously moving target or source of information indicates that OSINT is also continuously moving and capturing the pieces of interest that are more involved than a static, look-up of information on the web for one time like reading a newspaper one time. OSINT is now part of a different and dramatically more powerful world, just like movies versus static images.

Scanning web services from millions of sources, their comparison and interaction such as comparing Fox News to CNN International for the same stories, as well as stories exclusive to one network obviously yield dramatically more information than just reading the web for a story about an event. Even more striking is using OSINT from other countries, such as Mexico where a newspaper story about an event such as the recent killing of an ICE agent is dramatically different in Mexico than US media coverage—generally with detailed pictures, insights, and especially comments beneath the story. Within minutes of most major

events, stories and pictures will appear in Mexican electronic media such as *El Universal*, *La Reforma*, *Diario del Narco*, and *AFN Tijuana*.

Normally these electronic media stories are far ahead of US stories in both detail and depth of coverage especially with detailed pictures. Doing OSINT only in English practically guarantees that the best OSINT is not discovered or captured. Setting a web search engine to only collect data in English, as most browsers in the US are likely set, negates finding the robust web services that can assist with OSINT acquisition.

Similarly, only looking at text also negates most of what can be fruitfully found. Continuing the preceding example from Mexico, if only textual information is sought, then enormous resources in Social Networking sites like Twitter and Facebook are invisible. OSINT clearly doesn’t mean going to the Twitter and Facebook corporate offices with a warrant and getting all their data. It means logging on to Twitter and Facebook pages and discovering what is there, including the network of who is talking to whom, when, in what order, and in response to what? Twitter, as an example, can be a profound source of information by simply joining (or Following) some group or organization such as iRevolution. This group has done absolutely remarkable things to demonstrate the effectiveness of Social Media for disrupting major events like the G-20 conferences through “social activism” efforts. Joining and thus having access to the data in an open-source way simply takes knowing that the group exists. Yet, probably most such “digital activism” or “digital anarchy” groups are invisible to most of the IC simply because they don’t know what to look for and how to find the networks.

Obvious examples are the recent use of these networks for demonstrations in Egypt, Tunisia, Libya, Iraq, Yemen, China, and many other countries. For instance, in Egypt a

---

7 For more information, see *El Universal*, http://www.eluniversal.com.mx.
8 For more information, see *La Reforma*, http://www.reforma.com.
9 For more information, see *Diario del Narco*, http://www.diariodelnarco.com.
10 For more information, see *AFN Tijuana*, http://www.afntijuana.info.
11 For more information, see *iRevolution*, http://www.irevolution.org.
12 For more information, see *Egypt’s Uprising*, http://www.facebook.com/ElShaheeed.
Facebook page helped ignite the Egypt uprising, and the Twitter site (Jan25)\(^1\) done by a UCLA graduate student defeated the “unplugging” of the Internet in Egypt by using cell phones to send the information that was then tweeted to the world from outside Egypt. Also, there was the site that instigated massive demonstrations in Iraq\(^2\) with an advocacy that started almost a month before the Egyptian demonstrations on February 25. Hundreds of similar pages were there, most in Arabic and most with profound insight, yet the US State Department talked about being completely “blind sided” by Egypt. Having the US IC beaten by people openly using tools that millions to tens of millions were reading before they were in the open source world is a profound indication that OSINT is not all that it could be.

Another example from Mexico is the use of YouTube, which is much more widely used per capita in Mexico than in the US. Cartel videos bragging about their killings and showing the confessions of opponents naming hundreds of police and government leaders on the take, then the killing of the people, are incredible OSINT forms of information. However, the search in Spanish for such YouTube media before they are taken off the web is completely being missed by IC, that are trying to impact the world. Similarly, mining the comments below narco stories—where the comments in their jargon and slang must be translated by a native speaker very familiar with that world—is largely ignored. Yet, like most such comments in the US, people say far more than they should at times, and the “players” are often very visible to someone familiar with the concerns.

Even more revealing are the *narcocorrido*\(^3\) songs. Listening to the lyrics and knowing the players, then connecting the OSINT dots is likely not often done because it entails collecting lyrics, going to performances, and understanding the complex interplay of lyrics, events, and taunting that is going on with the songs. It is all open source and in the “entertainment world” but largely unknown and ignored because it isn’t the same traditional text. Establishing web services that could help bring in such information from bands, lyricists, YouTube executions, newspaper comments, Twitter accounts, Facebook accounts, etc.

---

\(^1\) For more information, see Jan25, [http://twitter.com/jan25voices](http://twitter.com/jan25voices).

\(^2\) For more information, see Iraq Demonstrations, [http://www.kitabat.com/thawra-25.htm](http://www.kitabat.com/thawra-25.htm).

and dozens of other open information could be built and may already be in the IC, but are certainly not deployed into normal Law Enforcement. Figure 1 is lyrics of an excerpt of a song that best exemplifies my statement on Narcocorridos; if DHS, IC and its DP evaluated its context, it could be used for intelligence purposes.

!["EL CORRIDO DE TONY TORMENTA"

Lyrics to an excerpt of "The Ballad of Tony Tormenta" by Chuy Quintanilla

Todo esta bien controlado se ñores
hagan conciencia; ahora que ya
tomo el mando el señor Tony tormenta

Everything is well under control, be aware se ñores, Now that Mr. Tony Tormenta is in charge

Todo los grandes se juntan para
escoger al mejor pero si se
necesita tambien sabe usar las
armas

All the big shots get together to
choose the best man, But if it's
necessary they also know how to
use their weapons

Figure 1. Lyrics to “El Corrido de Tony Tormenta.”

This is especially true since many government groups cannot log into Twitter, Facebook, YouTube, or similar sites because of the network rules for their agencies. In essence, we have conceded the vast majority of near real-time data to the “bad guys” and will get further and further behind as the speed of electronic, open source communications accelerates. Near the end of 2010, it was estimated that there were over 90 million Twitter tweets per day and over 5 billion Facebook messages per day. Adding dozens of other open-source communication tools to this, billions of pieces of OSINT are available to be analyzed daily, which in most government agencies is impossible.

Rather than giving up in despair, a simple comparison can be done in looking at web services such as Google, Bing, Yahoo, Amazon, and many more. This is done every day and is a money-making business. An example is the credit card monitoring company Fair-Isaacs which analyzes billions of transactions a day and stops staggering amounts of fraud with web services available to them online. Similar efforts could be designed within the IC, with the web services then being deliverable like Intellipedia out to many different clients so that the services could be amortized across many agencies. Working through the Office of the Director of National Intelligence (ODNI), such web services could be focused on efforts like counter-cartel efforts for US and Mexico, Middle East unrest, Humanitarian Assistance Disaster Relief (HADR), and eDiplomacy (State Department). DHS applications could be a subset of this and could enable DHS to use the web services rather than developing the expertise to develop and serve them. But without a decision to go this direction, ever expanding efforts by the world via Social Networking sites, YouTube, online newspapers, and non-English websites will mean that the US is doomed to failure in the IC’s ability to rapidly gather, digest, and provide actionable intelligence to decision makers.

According to the Director of National Intelligence (DNI), “The DNI is committed to ensuring the active and efficient use of open source intelligence, information, and analysis by the IC through the establishment and maintenance of an effective, reliable, and collaborative capability that provides maximum availability of open source information to all consumers, optimizes resource utilization, and establishes effective burden sharing.” How is this actually being done? How much further behind is the US because the materials for OSINT have exploded? What tools as web services are really there for the IC, for Law Enforcement or groups such as DHS? What are the simple tools that terrorists and cartels are using to blow past US Law Enforcement and IC?

DHS very much has the opportunity to make a major positive impact on the usefulness and impact of OSINT on behalf of the nation by seeking to develop, deploy, and optimize tools in the OSINT world using the opportunities and challenges that massive new

---

data sources present. Opportunities are available with the contextualization of data such as by Twitter and Facebook where information is accompanied by real relationships and where someone generally joined or “opted in” rather than simply anonymously doing a Google search. Using web services to process millions to billions of pieces of information per day and to build data infrastructure and workflows to assist with the broad range of DHS responsibilities appear to be an extraordinary opportunity at this time (end of 2010, beginning of 2011). This thesis is directed toward that goal by looking at the historical context of OSINT and trying to see how this same construct might be enhanced by DHS and other agencies to better accomplish the goals and responsibilities that they carry.
CHAPTER 2

OSINT IS NOT NEW IN THE INTELLIGENCE COMMUNITIES (IC)

FBIS began as the Foreign Broadcast Monitoring Service [see Figure 2] when in February 1941 President Roosevelt directed the Secretary of the Treasury to allocate $150,000 for recording, translating, transcribing, and analyzing certain radio programs broadcast from foreign transmitters--originally the Axis Powers' shortwave propaganda broadcasts. FBIS' overseas monitoring operation, which centered on radio and press agency transmissions, grew in response to the Cold War and worldwide U.S. interests. Its mission was significantly enhanced in 1967 when it assumed responsibility for foreign press exploitation, and its focus became not just foreign broadcasts, but all foreign mass media, broadcast and print alike. In subsequent years expansion of that mission has been driven by media dissemination technologies, which has meant new efforts required by, among other things, increased television and satellite broadcasting. Most recently, our mission has been expanded to include commercial and government public data base and gray literature. Our focus across all of these media types however remains foreign collection, principally in foreign languages.18

OSINT is not new to the IC; the CIA among other agencies has had a stake in OSINT for over 50 years.19 According to the Weapons of Mass Destruction (WMD) Commission Report,20 one of the primary recommendations, post 9/11, was to establish an Open Source Center within the CIA. This recommendation truly indicated at the time (and presently) to the IC the lack of resources to manage public “open source information in order to support their missions due to worldwide increase in media content and diffusion of communication technologies.”21

---

19 Wallner, “Myths and Realities,” 21.
Figure 2. The Foreign Broadcast Monitoring Service (FBMS) listening to broadcast information from countries.

Seeing that the CIA at the time had more than 50 years of experience in evaluating OSINT, the DNI agency decided that the most appropriate objective was to establish a center from its existing Foreign Broadcast Information Service agency within the CIA; better known now as the Open Source Center. The FBIS has evaluated OSINT and it has been successful on meeting its objectives; its content of collected information has been utilized to gauge local reaction to events, to estimate future events, identify rhetoric patterns and broadcasting schedules to support intelligence analysis; which DHS can very well mimic at the domestic level. The importance of FBIS to the modern intelligence world was summed up in a *Washington Times* article in 2001: “So much of what the CIA learns is collected from

---

22 For more information, see Open Source Center, https://www.opensource.gov.

newspaper clippings that the director of the agency ought to be called the Pastemaster General.”

In his article, “Open Sources and the Intelligence Community: Myths and Realities,” Wallner elaborates on the criticisms of what the Intelligence Communities have done to upgrade the Cold War era mentality to a more adaptive force by suggesting that for some time the FBIS was doing its job on examining open source data at the CIA level for over 50 years. Wallner concludes by suggesting that the reality is that the IC at all levels have been unable to keep up with the pace of qualitative and quantitative open source information since the 1980s mainly due to traditional methods of acquiring information, less cost effective measures, and more risky intelligence methods that most agents thrive on, better known as the spy game. Even with this in mind, Wallner believes that OSINT information contributes much more to the Intelligence Community, as the following states:

Science and Technology analysis estimate that 70 to 80 percent of their knowledge base is derived from open information. Economic and political users judge that open-source material gives them some 30 to 40 percent of their fundamental data needs, while military customers assess that about 25 percent of their needs are met through open-source information…Arms-proliferation, terrorism, and counternarcotics intelligence elements are busy incorporating a surprising amount of open data into operations, as well.

There is no question that the FBIS has paved the way for other agencies such as the CIA to continue the tradition of evaluating OSINT through the current Open Source Center, interestingly enough the CIA is in charge of foreign intelligence. However efficient the Open Source Center has been since its incarnation in 2005, it still denies DHS the capabilities to harvest its own domestic OSINT. This doesn’t mean, we shouldn’t learn from our predecessors, but should acknowledge the need to conceptualized DHS’s very own Domestic Open Source Center and the enormous power OSINT has to offer.

DHS needs to incorporate its own Open Source Center in order to facilitate the needs of its organization domestically to fight terrorism and criminal activity in the 21st century (as

24 Ibid.
26 Ibid., 20.
FBIS has shown us). By utilizing OSINT as a main frame for gathering information and evaluating all sources domestically, specifically Internet information, DHS can thwart the purposes of criminals and terrorists in the US, reduce our vulnerability, and enhance recovery measures that impact the freedoms of citizens. It’s unfortunate that the general public has lost much of its trust in the government originally meant to help them. Many citizens are asking themselves, “What is the government really doing for us?” “Can they really protect us?” “And at what cost in personal freedoms and/or actual dollars?”

Many citizens today are again questioning the Intelligence Community’s activities, including its competence and professionalism after learning about serious intelligence failures leading up the 9/11 attacks in September 2001, and prior to the Iraq war in March 2003. These very public failures, combined with controversial practices such as renditions and detentions of terrorism suspects, warrant-less searches and widespread eavesdropping against U.S. citizens, have sparked greater demand for transparency in Intelligence activities. ²⁸

DHS needs to revamp the intelligence cycle and incorporate practical and instrumental OSINT technologies (web services) as part of their intelligence objectives. DHS was developed shortly after 9/11, ²⁹ according to the Homeland Security Act of 2002. ³⁰ DHS’s current mission is to: “prevent terrorist attacks within the United States; reduce America’s vulnerability to terrorism; and minimize the damage and recover from attacks that do occur.” ³¹ If the objective is to prevent and reduce terrorist attacks and recover from them, as suggested, then DHS should look no further and incorporate OSINT and its web services as a tool to minimize the threat of terrorism in the US. Considering that the general population is becoming more adaptive to the Internet, including criminals and terrorists, this should prompt DHS to take more adaptive approaches towards the Internet if its objective is to proactively minimize threats to the nation.


The Subcommittee on Intelligence, Information Sharing, and Terrorism Risk Assessment Report: “Using Open-Source Information Effectively” examines the need to use open-source information effectively in order to strengthen DHS information analysis and intelligence production responsibilities. Unfortunately, “to date, despite extensive support from the Director of National Intelligence (DNI) and a statutory imperative to act, DHS has yet to stand up to a robust open source program to share information with law enforcement and other appropriate domestic partners,” as it should do, considering the domestic intelligence role it plays in the US. “Since the 9/11 terrorist attacks, Congress has not only focused considerable attention on how intelligence is collected, analyzed, and disseminated in order to protect the homeland against terrorism, but also what should such intelligence encompass.” In retrospect, the creation of the Homeland Security Open Source Information Enhancement Act of 2008 (H.R. 3815) was developed. Clearly, this stipulated that DHS had not conceptualized or even emphasized enough attention on OSINT in order to combat the threat of terrorism in the US.

Congress has illustrated greater emphasis on Internet intelligence gathering, among other concerns; unfortunately, due to DHS’s reorganization and establishment of organizations, it has been unable to capitalize on the production of such analytical products, as well as these two other functions: (1) All open source products should be readily available to the public sector (this includes “Federal, State, local, and tribal law enforcement” organizations), and foreign allies; and (2) DHS has a stake in “providing open source products to consumers” as the federal guidelines have stipulated and as an entity of domestic intelligence.


And more importantly, Congress has stated that it’s the responsibility of the Secretary of DHS to establish an open source collection, analysis and dissemination program which will incorporate these elements: (1) analyzing “risk and vulnerabilities to the nation’s critical infrastructure;” (2) analyzing terrorist or criminal tactics and techniques; (3) analyzing “news and developments related to foreign terrorist” or criminal activity; and (4) analyzing better resources for product development in open source intelligence gathering.\textsuperscript{36}

Therefore, the technical meaning of OSINT (end product) is: “information that is publicly available and that can be used and disseminated in a timely manner to an appropriate audience for the purposes of addressing a specific homeland requirement.”\textsuperscript{37}

In the 21st Century, detailed data about almost anything or anyone is accessible on the Internet. Although much of this information is for social or commercial use, some of it may be helpful to terrorists as they plot attacks against the homeland. Put simply, there is a world of publicly available information online that terrorists might exploit to cause death and extreme damage to the Nation’s critical infrastructure. Open source information is a valuable source of data about terrorists themselves that must be fully integrated into the intelligence cycle to ensure that policymakers are fully and completely informed about threats. It therefore should be viewed not only as a supplement to classified data but also as a potential source of valuable intelligence that can make the Nation safer. Open source has the added benefit of being open, unclassified information that can be shared rapidly with State, local, and tribal law enforcement and other first preventers without the need for security clearances.\textsuperscript{38}

\textsuperscript{36} Ibid.
\textsuperscript{37} Ibid.
\textsuperscript{38} Ibid., 3.
CHAPTER 3

OPEN SOURCE INTELLIGENCE

As we are beginning to learn, OSINT can efficiently shape how DHS combats the new generation of technological terrorists operating through the Internet that are trying to spread radicalism, plot terrorist activities, and diffuse bigoted rhetoric across the Internet. “Public available information is the foundation upon which all intelligence operations build to develop all-source intelligence that supports situation understanding and decisive action. The availability, depth, and range of publicly available information enable intelligence organizations to satisfy intelligence requirements without the use of specialized human or technical means of collections.”39 This is further stipulated in the DNI report, The National Intelligence Strategy of the United States of America:

> Intelligence Community faces an explosive growth in type and volume of data, along with an exponential increase in the speed and power of processing capabilities. Threats to our networks and the integrity of our information have proliferated. Our partners and users increasingly expect us to discover, access, analyze, and disseminate intelligence information in compressed time frames. We have the responsibility to share information, while protecting sources and methods and respecting the privacy and rights of U.S. citizens.40

OSINT information is not classified and is available to the public. As noted previously, OSINT material derives from newspapers, the Internet, book journals, periodicals, radio, television, commercial databases, and gray literature like think tank studies, symposium reports, academic research, corporate brochures and reports that are free of proprietary or copyright restrictions.41 Hence the reason to look at the Internet, including Social Media outlets such as Facebook, Twitter, and dozens of others that are becoming prominent pools of open source data, largely available to anyone, including DHS. Since

---


41 Wallner, “Myths and Realities,” 19.
DHS’s historical incarnation in the 21st century and reorganization, Congress has emphasized the need for DHS to be more proactive in creating intelligence products, and implementing OSINT as part of their intelligence strategy. As it stands, law enforcement and other units of intelligence have reiterated the need to use OSINT more effectively, and expect DHS to lead them in acquiring more productive intelligence. A report by the House of Representatives Committee on Homeland Security, “Giving a Voice to Open Source Stakeholders: A Survey of State, Local, & Tribal Law Enforcement,” reported findings on the DHS:

82% of respondents reported that they collect and analyze open source information, with a majority expressing a desire to raise their situational awareness of “all hazards” and an interest in receiving DHS open source products providing that information. Only 50% of respondents, however, reported that DHS open source products met that need. Moreover, 60% of respondents reported that in order to improve matters, DHS needs to establish a robust training program in addition to producing open source intelligence products with actionable recommendations.42

It is obvious to many members of Congress and IC that all OSINT information should be considered potential sources of intelligence. Whether it would detract from old methods of analyzing intelligence is out of the question, only those acquainted with the Cold War Era mentality would believe this; however, the enemy has shown us that they are more resilient and radical than before. With this in mind, “the Intelligence Community should devote more resources to exploiting information that is available publicly, collecting materials that do not circulate widely, and assigning a significant analytical effort to monitoring OSINT.”43

The end result is that, “open-source information, when properly assembled and analyzed, can provide some of the most strategic, tactical and operational data imaginable in order to obtain an ever-evolving, near real-time picture of terrorists’ plans.”44 DHS’s commitment since 9/11 has been to install an OSINT program; however, they have been missing the mark completely within their own organization and have not been providing their domestic partners with the necessary products to combat the threat of terrorism, as has been

42 House Committee on Homeland Security, Giving a Voice to Open Source Stakeholders, 3.
44 House Subcommittee on Intelligence, Using Open-Source Information Effectively, 4.
pointed out in this thesis. As of September 2008, *H.R. 3815* directed DHS “to establish a program to collect and analyze open source (publicly available) information and disseminate reports and other products based on information from DHS. CBO expects that the department would hire about 10 people to carry out the activities required by the act and estimates that the additional staff would cost about $1 million annually over the 2009-2013 period, assuming the availability of appropriated funds.” What this suggests once again is that the actionable recommendation from Congress, *H.R. 3815*, and the National Intelligence Strategy has yet to be fulfilled or implemented by DHS to a substantial degree.

In order to truly benefit from OSINT, DHS needs to fully understand what are the benefits and liabilities of OSINT and find the right tools to do the job. It doesn’t matter how much funding is available or has been accounted for, if you don’t know how to conceptualized product availability (web services) and data capacity at the same time, then you will not understand OSINT capabilities. Once DHS is able to understand this concept, DHS will be able to see the greater picture of what OSINT really has to offer. In essence, determining what OISNT is, and how pre-existing web services accentuate intelligence and how it can be sharable among other intelligence partners is what is important. In his paper, “Terror on the Internet,” Weimann further elaborates on the notion of OSINT information that is vital and relevant to DHS, IC and its DP:

> We know that terrorist organizations are increasingly resorting to the Internet to disseminate their views to a wider public, coming to the realization that establishing their presence in cyberspace is nearly just as critical to their long-term success as any military triumph or act of sabotage. Terrorist groups themselves can maintain webpages to “advertise” their ideology, disseminate propaganda and recruit supporters. It is the first time that they can easily reach the public directly and make their existence known in an international scale. The web allows an uncensored and unfiltered version of events to be broadcast worldwide. Chat rooms, websites, and bulletin boards are largely uncontrolled, with few filters in place.46

---


In his report, “Open Source Intelligence: A Strategic Enabler of National Security,” Pallaris explores the *benefits* of OSINT and notes that the immediate benefit associated with OSINT is “cost”, as well as “collecting information via classified means.” Further illustrating that OSINT is shareable among all facets of intelligence, Pallaris gives examples: “OSINT collected under ethical means can be used in legal proceedings without risking exposure to sensitive intelligence assets [e.g., informants] . . . OSINT can be accessed, exploited, and shared around the clock . . . [any] intelligence gathered from authoritative public sources can be used to inform the public of serious threats to national security;” and Pallaris suggests that it “provides context and awareness that is critical to understanding the global security agenda.”

The *limitations* of OSINT are quite understandable, but workable with a knowledge of OSINT’s true harnessing power. According to Pallaris, “It is unlikely to offer 100%” information to intelligence communities due to various factors: “information overload ... filtering the ‘signals’ from the ‘noise’ is becoming increasingly difficult; ... [and] technology vendors” that have the latest and greatest software have holes in programming; and technology will still not be able to replace the analytical efforts of “human beings.”

Various academic sources affirm the unquestionable benefits of OSINT’s power are immense, with little negative contention, as shown in Table 1. It is clearly seen that the benefits outweigh the liabilities by a great margin. The determining factor of what makes OSINT valuable and important to DHS is that consumers (people using the internet) are the primary hosts of technology consumption (making the Internet that much more valuable to DHS); therefore, there is no way one can ignore the potential harvesting power that OSINT has to offer DHS, via the Internet. With a thorough look at these benefits and liabilities, it may be concluded that OSINT has an unrepresented value in which DHS, IC and their DP can appreciate, acknowledge and utilize in order to preempt and combat terrorism via the Internet.

---

### Table 1. OSINT – Benefits vs. Liabilities

<table>
<thead>
<tr>
<th>Open Source Intelligence (OSINT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
</tr>
<tr>
<td>it’s fast, flexible, dynamic and cheap to access&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>approximately 80 percent of intelligence information derives from open source&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>it is communicable, sharable, and trust-creating for NATO and alliances&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>that have limited and restricted access to intelligence information&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>identifies and mitigates risk at strategic, operational, tactical, and technical levels&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>spans quick and dirty evaluation to in-depth analysis of current issues at hand (on-going)&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>contextualizes intelligence on current and historical means, re-directing other resources when needed&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>contributes to all-source collection process to the intelligence community, freeing-up other resources&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>ensures decision makers at all levels that all sources of information have been evaluated&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>collected under ethical means it can be used in legal proceedings without risking exposure to sensitive information or intelligence assets (like informants)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>intelligence gathered from public sources can be used to inform the public of serious threats to the nation&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

*(table continues)*
Table 1. (continued)

<table>
<thead>
<tr>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>it’s unlikely to offer 100% information to Intelligence Communities, without having to cross-reference</td>
</tr>
<tr>
<td>software needs to be intuitive and learnable and robust to capture data</td>
</tr>
<tr>
<td>information overload (what is considered good and bad information)</td>
</tr>
<tr>
<td>trusting vendors (private organizations) and their software developers to capture needed information</td>
</tr>
<tr>
<td>technology will not be able to capture or replace analytical efforts of human beings</td>
</tr>
<tr>
<td>language barriers among analysis and IT specialists evaluating OSINT</td>
</tr>
<tr>
<td>OSINT is not valuable if there's not any qualitative information to analyze</td>
</tr>
</tbody>
</table>

---

b Wallner, “Myths and Realities,” 20.

It has come to my attention that OSINT is valuable and needed at DHS; there is no doubt about that. The information presented has proven what OSINT can and will offer, as well as the restrictions that DHS could face when inscribing a good OSINT center, training program, and so forth. For further clarification, “Open source intelligence is the outer pieces of the jigsaw puzzle, without which one can neither begin nor complete the puzzle...open source intelligence is the critical foundation for all-source intelligence products, but it cannot ever replace the totality of the all-source effort.”⁴⁸

It’s obvious that every benefit has its liability (as we can see in Table 1). In the analysis below, we explore some relatively new concepts, as they add significant value to

---

DHS’s apprehension in utilizing OSINT more effectively and efficiently. Moreover, it’s pertinent to illustrate these concepts in advance since a thorough knowledge of your objective can very well determine your outcome, even though liabilities exist. One can overcome apprehension and retain objectivity through conceptualization and trial and error; this DHS can and will accomplish if terrorism prevention, reduction and recovery is part of its mission statement. The following analysis is based on my own conclusions set forth by the compilation of benefits and liabilities.

- **Negligence to illustrate whether OSINT can really deter violence, terrorism or crime in the US:** As illustrated, foreign information has been analyzed by organizations like the CIA for over 50 years; however DHS is still new in collecting OSINT according to the “Homeland Security Open Source Information Enhancement Act of 2008”, and “Using Open-Source Information Effectively” making it difficult to see what is the true value or measure of success in DHS. So far there has not been any cases which I have come across that confirm the use of OSINT to deter or prevent a crime or act of terrorism through DHS, if they’re limited by their availability to the public. I do understand the level of confidentiality in government issued reports. It goes to show that qualitative information needs to be significant or vital in order for OSINT to work, making the task much harder for DHS to evaluate OSINT.

- **Appears that OSINT is institutionalized and usable only after a crime or act of terrorism has occurred:** It’s quite simple; much of the information presented before you illustrates the obvious but neglects the deterrence factor. Most criminals or terrorists are discreet about their crimes, to some degree. Most don’t want to be caught, and those that do want to be caught have a psychosis revelation of why they did it. Second, most highly trained terrorists so far have taken quite a few steps to conceal their plots after the act of terrorism has occurred. Most often they take credit for their crimes through video or rhetoric in some Internet page, as to why they did it. So, OSINT really is utilized after the fact not before which absolutely contradicts the efforts of Intelligence Communities and DHS, utilizing OSINT to some extent.

- **Pieces of the puzzle don’t come before but after a crime has been committed:** In “Open Source Intelligence (OSINT): Issues for Congress,” Best and Cumming, cite Joseph Nye’s observation as to the dominant perspective of the Intelligence Community, “Open source intelligence is the outer pieces of the jigsaw puzzle [that come into play as an] all-source intelligence”\(^49\) package; miss one piece and you have an incomplete puzzle in front of you until disaster strikes. More or less this puzzle piece comes with severely superficial information with hardly any clear affirmation or mention as to when an event or crime will be committed, with the need to cross-reference much of the information.

\(^{49}\) Ibid.
• As much as OSINT contributes (80% is the figure given by organizations like the CIA that do foreign surveillance) to the Intelligence Communities, it's still dubious in preventing catastrophic events at the domestic level: The events of 9/11 proved this. We had our fair share of information (CIA which has had over 50 years of OSINT and the FBI which examines domestic issues for the homeland failed to communicate with one another), and nobody moved on this issue, until it was too late. Leading to the obvious question, would DHS, the FBI, the DNI and the CIA be able to see the big picture when other terrorist plots are being formulated and sources like OSINT are available to be examined and analyzed?

When trying to understand the benefits and liabilities of OSINT, you conclude that it’s not whether a table shows more or less benefits and or liabilities (this is only a visual aid on OSINT). It is how you determine the flow of information to be valuable and resourceful using OSINT and the web services that are available to you to enhance its capabilities (collection and utilization). Even as I express my thoughts on DHS’s uneasiness to utilize OSINT more efficiently and resourcefully, it is clear that OSINT still has so much more to offer DHS, IC and its DP than elaborated on here.

The end result is to find new ways to preempt, deter and strike back at criminals and terrorists alike, in order to secure the US; OSINT is the best preventative tool DHS can access and utilize. The Internet has shown the world where it’s moving and where it can go, and that is why DHS needs to be proactive and reactive at the same time, keeping in mind the endless possibilities the Internet has to offer. Thomas L. Friedman best illustrates this concept, by discussing the power of uploading, also known to many as blogging (where many terrorists are uploading their own rhetoric and bigotry of the US, material that DHS could analyze through OSINT and greatly get the upper hand on preventative measures):

Soon after the community-developed software movement gained momentum, we witnessed the emergence of another bottom-up, self organized form of uploading: blogging. I see this most vividly in my own profession, journalism, where bloggers, one-person online commentators, who often link to one another depending on their ideology, have created a kind of open-source newsroom. A blog is your own personal virtual soapbox, where you can get up every morning and, in the form of a column or a newsletter or just a screed, tell the world what you think about any subject, upload that content onto your own Web site, and then wait for the world to come check it out. If others like it, they will link to your blog from their blog or to other kinds of content, like online news article or commentaries. I now read bloggers (the term comes from “Weblog”) as part of my daily information-gathering routine. In an article about how a tiny group of relatively obscure news bloggers helped to blow the whistle that exposed the bogus documents used by CBS News’s Dan Rather in his infamous report about
President George W. Bush’s Air National Guard service, Howard Kurtz of The Washington Post wrote (September 20, 2004), “It was like throwing a match on kerosene-soaked wood. The ensuring blaze ripped through the media establishment as previously obscure bloggers managed to put the network of Murrow and Cronkite firmly on the defensive. The secret, says [Web designer and blogger] Charles Johnson, is ‘open-source intelligence gathering.’ Meaning: ‘We’ve got a huge pool of highly motivated people who got out there and use tools to find stuff. We’ve got an army of citizen journalists out there.’”

The immense allure of blogging and other informal Social Media have presented many opportunities and challenges for DHS, IC and its DP to appropriately utilize OSINT and Social Media/Networks/Blogs/Search Engines and sites. Social Media/Networks/Blogs/Service Engines like Facebook, Twitter, MySpace, LinkedIn, Hi5, V Kontakte, Orkut, Tagged, QQ, Friendster, Xianonei, Mixi, Maktoob, YouTube, Flickr, Google, Yahoo, and Bing are all prominent un-tap web service resources that DHS, IC and its DP have not conceptually made part of the OSINT cycle.

In many complex ways these tools can “reveal personal communications,” “establish motives”, and display “personal relationships” as never before possible. They can also provide “location information,” “prove and disprove alibis”, and help enable crime, criminal enterprise, and terrorist organizations. Social-networking sites generally allow users to “create personal profiles,” “write status updates or blog entries,” “post photographs-videos-audio clips,” “send and receive private messages,” and “link to the pages of others” (i.e., “friends”).

This information is generally transient, as the writer can add or subtract information at will. The information also has no independent check of its validity—it can be completely false or partially false, so that the information is simply information with no guarantee of anything being true or false. For the most part IC, DHS and its domestic partners can obtain information from these sites just as anyone else would by viewing the information that is open to public. Simply “friending,” “following,” “liking,” or other ways of joining open networks are very simple ways to acquire open-source information, just like reading a

---


newspaper. Additionally under appropriate legal guidelines, DHS and Law Enforcement groups can use the Electronic Communications Privacy Act (ECPA) to legally obtain information from providers for investigations within the mandated responsibilities of DHS.52

What this illustrates is that Social Media/Networks/Blogs/Search Engines are fast becoming part of many people’s “identity” and have become social outlets for massive amounts of information available about people in general but also about many terrorists who also choose to post information about themselves and their causes online. DHS can benefit from this massive amount of information in its abilities to protect the nation, but only if it seeks out this information and appropriately uses it, including connecting the dots by linking the analysis of pieces of information over periods of time.

Figure 353 shows a visual entitled “The World Map of Social Networks” which illustrates the extraordinarily global spread of Social Networks as of December 2010 that now involve billions of messages a day linking relationships among people. These relationships and their content could provide DHS and the IC with extraordinary information, but only if it is sought after, recorded, and visualized using such software as Semantica by Semantic Research where complex link analysis can be done. Without seeking and recording the Social Networking, or using the web services where it is recorded and available, the DHS is essentially gifting this massive capability to the “bad guys” and is missing the largest and fastest growing amount of information in the Open-Source (OSINT) world.

52 Ibid.
Figure 3. World map of social networks.
CHAPTER 4

HARNESSING THE POWER OF THE INTERNET
AND ITS WEB SERVICES

Much can be said about the power of the Internet; it helps IC the Intelligence Community connect with one another, enables users to interact in ways unimaginable decades ago and helps in acquiring OSINT. As the DHS decides in which direction to best pursue OSINT, one wonders how DHS can best acquire and use the Internet to harness its power and use it to gather intelligence.

As you can see in Figure 4, this chart presents a listing of primary open source media (Internet sites) that well-trained collectors can detect and monitor. Internet sites of interest may provide enemy intentions, capabilities and activities.\(^ {54}\) The importance of this data is to once again illustrate what OSINT has to offer DHS and the hub of information that can be converted into intelligence. Additionally, Pallaris and Costigan point to the frame of mind DHS needs to have when conceptualizing OSINT, in order to be proactive in utilizing Internet web services:

A first step in this direction would be to harness the benefits offered by open source learning environments. The last decade has seen the evolution of the Internet from an information-based publishing platform to a knowledge-based learning platform. It is now possible for any individual to go online and contribute what they know to a common pool of knowledge. This knowledge can be reviewed, revised, repackaged and redistributed in any number of forms – a news article, a research paper, a learning module or an anecdote.\(^ {55}\)

In “Open-Source Spying,” Clive Thompson illustrates the obsolete system (called Intelink) the Intelligence Communities were using to communicate and gather intelligence with one another. Outdated and un-resourceful to the IC, they soon realized that Intelink was

\(^{54}\) Department of the Army, *Open Source Intelligence (Field Manual Interim No. 2-22.9)* (Washington, DC: Department of the Army, December 5, 2006), chap. 2 (Fundamentals), p. 5

27

Figure 4. Internet site categories – DHS can analyze for OSINT.

no longer providing the necessary functions to connect with one another, let alone provide reasonable intelligence. In the interim of learning new ways to communicate intelligence to one another, IC devised a Wiki base technology utilizing an Internet web service, called Wikipedia. The outcome came as no surprise, a riveting new intranet system was developed called Intellipedia which connected Intelligence Communities to one another in a more dramatic way.

Intellipedia became and is a source of connectivity and power almost unimaginable decades ago, using an adaptive web service provided by the Internet. Intellipedia is an intranet technology that allows IC to connect with one another and update information like Wikipedia, with one slight difference. The author who updates this information is not anonymous, so information is current, well researched and pertinent to each other’s case.56 Authors Pallaris and Costigan praise this Wiki Base web service:

---

This learningscape—currently best exemplified by the Wikipedia project—is a flexible and extensible learner-centric environment that provides tools, resources and community support around one or more subjects. The success of Wikipedia and its philosophy hasn’t been lost on the US intelligence community; in October 2006 it announced the launch of Intellipedia, a wiki-based platform for greater knowledge sharing, data synthesis, and collaborative exchange. By using wiki technology, the US intelligence community is hoping to leverage its intellectual assets in a collaborative way for the production of national intelligence estimates and other key reports.57

Furthermore, if the IC has already zeroed in on a Wiki base web service like Intellipedia and it seems to be working; why wouldn’t it zero in another form of a web service, like Google? Google, rated as a premier search engine among consumers for its usage, apparatuses and web content, has made small headlines in the field of information gathering (collection and utilization). Verne Kopytoff suggests that Google has been providing important collection and utilization apparatuses:

When the nation’s intelligence agencies wanted a computer network to better share information about everything from al Qaeda to North Korea, they turned to a big name in the technology industry to supply some of the equipment: Google Inc. Many of the contracts are for search appliances-servers for storing and searching internal documents. Agencies can use the devices to create their own mini-Googles on intranets made up entirely of government data. Spy agencies are using Google equipment as the backbone of Intellipedia, a network aimed at helping agents share intelligence. Rather than hoarding information, spies and analysts are being encouraged to post what they learn on a secure online forum where colleagues can read it and add comments.58

According to the American Customer Satisfaction Index (ACSI), search engines like Google, Alta Vista, Bing, Yahoo, MSN, AOL, and Ask rank among the most satisfying Internet portals and search engines to consumers. Among Internet news and information sites, Foxnews, USA Today, NYTimes, ABCNEWS, MSNBC, and CNN rank highest in customer satisfaction.59 How does this information apply to IC and DHS? Customer

---

57 Pallaris and Costigan, “Beyond the Age of Information,” 7-8.


satisfaction prompts usage, usage equates to information (web content traffic), information equates to intelligence (querying web content for radicalism and important rhetoric information valuable to DHS), making search engines and Internet news and information sites reliable mechanisms for information gathering, evaluation and intelligence. Table 2 illustrates the current customer satisfaction index, which DHS can analyze and infer that consumers are extremely drawn to the Internet, which makes the Internet that much more valuable to DHS. The index shows where people are going and not going, which is equally important for information.

Table 2. ACSI Internet Portals and Search Engines, News and Information, and Social Media Websites

| Internet News & Information | 97 | 98 | 99 | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | Previous Year % Change | First Year % Change |
|-----------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----------------------|---------------------|
| FOXNews.com                 | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 82                   | N/A                 |
| USATODAY.com                | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 77                   | 4.1                 |
| NYTimes.com                 | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 74                   | 4.1                 |
| ABCNEWS.com                | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 74                   | 5.8                 |
| MSNBC.com                  | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 73                   | 1.4                 |
| CNN.com                    | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 74                   | 2.8                 |
| All Others                  | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 73                   | 0.0                 |

| Internet Portals & Search Engines | 97 | 98 | 99 | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | Previous Year % Change | First Year % Change |
|-----------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----------------------|---------------------|
| AltaVista                        | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 77                   | -7.2                |
| All Others                       | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | N/A                  | N/A                 |
| Google                           | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 82                   | 5.1                 |
| Bing.com                         | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 80                   | -7.0                |
| Yahoo.com                        | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | N/A                  | N/A                 |
| MSN                               | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 76                   | -1.3                |
| AOL                              | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 74                   | 5.7                 |
| Ask.com                          | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 74                   | 1.4                 |

| Internet Social Media | 97 | 98 | 99 | 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | Previous Year % Change | First Year % Change |
|-----------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----------------------|---------------------|
| Wikipedia.org          | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 77                   | N/A                 |
| YouTube.com            | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 73                   | N/A                 |
| All Others             | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 72                   | N/A                 |
| Facebook.com           | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 64                   | N/A                 |
| MySpace.com            | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | NM | 63                   | N/A                 |
Currently, Google has been in the forefront of Internet technology, dominating the global market with its apparatuses, and currently it is the default page of many of its Internet consumers. As Professor Claes Fornell reports, “Portals and search engines lead all e-business categories with an ACSI score of 77, but declined 7% from a year ago. This decline was driven mostly by a drop for Google, which commands a large proportion of search engine and portal traffic. Despite Google’s 7% fall from an all-time high of 86 a year ago to a score of 80 this year, the dominant player in the search engine business still leads its largest competitors by a wide margin.”

**WEB TECHNOLOGIES: ALERTS, RSS FEEDS, WEB SLICES, AND READERS**

As more Intelligence Communities start relying on the Internet as a major source of gathering OSINT, the Intelligence Communities like DHS need to start exploring other Internet web technologies as we just witnessed with Intellipedia. There is much to be discovered. As I ponder, and ask what Internet technologies exist that can assist DHS in fighting terrorism or crime in the US, I look no further then to the most immediate technologies that come to mind: technologies like Alerts, RSS Feeds, Web Slices and simple Readers that can maintain the order of your information, ready to be analyzed by competent DHS analysts and molded into a processing intranet system to fit the current needs of DHS, as we have previously witnessed with Intellipedia.

Currently in the market there exist many kinds of Alerts provided by search engines like Google (noted in Figure 5), Yahoo (noted in Figure 6) and Bing (noted in Figure 7), to name a few. To illustrate, Alert’s monitor content that automatically notifies users when new content from news, web, blogs, video, and/or discussion groups matches a set of search terms selected by the user and stored by each Alert service provider. Examples of Alerts can range anywhere from finding out what is being said about a company or product; monitoring a developing news story; keeping up to date on a competitor or industry; getting the latest news

---

on a celebrity or sports team; and finding out what's being said about themselves (like DHS).\(^{61}\)

Web Slices on the other hand, allow certain portions of a web page to be subscribed to, and allow you to “follow” an area of a web page through a dedicated feed bookmarked in your desktop,\(^{62}\) as noted in Figure 8.

Real Simple Syndication (RSS) feeds allow people to see content without having to go to the website the html processor seeks publish updates, such as blog entries, news headlines, audio, and video in a standardized format for any person to read. For example RSS


Figure 6. Yahoo! Alerts interface.

Figure 7. Bing Alerts interface.
feeds are not just for blogs or social media sites; RSS can distribute many kinds of web content: news-press releases; educational articles; videos; destination information; tech support updates; product news; specials; announcements; the list is almost endless.\(^6^3\) RSS Feeds constantly checks your favorite news sites and blogs for new content; downloads any updates what it finds, and what a Reader like Google Reader does, it provides a user interface to monitor and read the feeds you have subscribed too,\(^6^4\) as noted in Figure 9.

DHS can very well take the above describe web services and utilizing them to enhance OSINT, they just have to find the best possible methods in implementing these web services for their benefit. One method, is utilizing all the search engine providers (like Google, Bing, Yahoo, Metacrawler, Dogpile and many more) and having all there information funnel through various hubs cross referencing each others data, for continuous alerts, web slices, RSS feeds and reading them in a specialized Reader (as I mention above), the idea would be to query the web's content in order to gather information, which in theory could be utilized for intelligence purposes.

I monitor a variety of Alerts and RSS Feeds for an average of 5 months to illustrate the potential extraction or query of web content from prominent search engines like Google,


Yahoo and Bing. Alert topics used were: crime and social networks, home-grown terrorism, international terrorism, Islamic propaganda in social networks, Islamic terrorist supporters, jihadist (video, blogs), jihadist and United States, jihadist propaganda, kill non-believers, Middle Eastern blogs, radicalization on the search engines, rise in militia groups in the US, terrorism blogs, terrorists in the US, and Allahu Akbar in the US. I had a total of 2596 alerts and 35 subscriptions over a period of five months illustrating the power of these web services.

I also monitored RSS feeds for about the same duration of time utilizing a Google Reader to monitor all the alerts being flagged. The alert topics where in excess of 1000 articles at a time (with at least 10 subscriptions), and as new articles came along they replaced the old articles that were read. These alerts were: Articles at Technorati (Blogosphere Website), All Africa News: Latest News, BBC News Africa World Edition, BBC News Arabic, China – The People’s Daily Online, Google Alerts – Islamic propaganda in social
networks, Pravda News (Russian News Website), RIA Novosti (Russian News Website), China.org (China Top News Website), and You Tube Matching Query: kill non-Muslims.

I independently kept web slices in my tool bar; however, I didn’t receive alerts, due to the nature of how web slices work. They’re directly linked to web sites that provide these web slices, and usually its content that is being updated on a particular web site. You’re most likely following an area of a web page that is of particular interest to you, bookmarked on your desktop.

What this illustrates is that if I can set up an alert monitoring system myself using these web services, imagine the capabilities that DHS would have to be able to build around using these web services (like building an intranet system using these technologies in one remote location, like a reader), to help assist them in fighting terrorism via the Internet. The importance of this is to showcase that these web services do work. Where they provide the necessary intelligence information platform is up to DHS to determined, but we can only imagine the possibilities. The end results are illustrated in Table 3. Following this table are examples of these alerts, shown on Figure 10 (Google Alert), Figure 11 (Yahoo Alert) and Figure 12 (Bing Alert). Following Figure 10 (Google Alert), Figure 11 (Yahoo Alert) and Figure 12 (Bing Alert) is my own brief analysis on web services and the shortfalls and gains of using these web services.

Table 3. Internet Web Services

<table>
<thead>
<tr>
<th>Web Services</th>
<th>Type</th>
<th>Months</th>
<th>Alerts/Subscriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Alerts</td>
<td>Varied</td>
<td>4 (2/11-6/11/10)</td>
<td>1622/16</td>
</tr>
<tr>
<td>Yahoo Alerts</td>
<td>Varied</td>
<td>5 (3/20-8/20/10)</td>
<td>305/10</td>
</tr>
<tr>
<td>Bing Alerts</td>
<td>Varied</td>
<td>5 (3/20-8/20/10)</td>
<td>669/9</td>
</tr>
<tr>
<td>RSS Feeds</td>
<td>Varied</td>
<td>5 (Avg.)</td>
<td>1000+/10</td>
</tr>
<tr>
<td>Google Reader</td>
<td>N/A</td>
<td>5 (Avg.)</td>
<td>N/A</td>
</tr>
<tr>
<td>Web Slice</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Figure 10. Google Alert.

Figure 11. Yahoo Alert.
WEB SERVICES - ANALYSIS

- Although Internet web services exist, better methods of exploitation need to be developed using existing technologies, to contribute to DHS deterrence of terrorism and criminal activity: This will signify that IT specialists need to be more competent in exploring these existing technologies and utilizing them to DHS benefit. And DHS will need to take a more proactive or take a more ad hoc approach in utilizing private institutions to develop these software technologies, as we saw with Google and Intellipedia. DHS will also need to acquire resources at the CIA and the DNI to best implement OSINT in utilizing these Internet web services.

- Alerts, Web Slices, and RSS Feeds are limited by Internet service providers (information), and following these technologies could be monotonous since they could be full of biased and untrustworthy information: Apparatus (technology) will never delineate its true measure of success, but what we can do is explore its capacity to do the best job possible. And information should be cross-referenced with other existing information to certify its accuracy and completeness. As I have noted previously, OSINT is but one facet of intelligence. The biggest weakness OSINT can have is not exploring its content and the existing technologies that can provide a much faster level of gathering information and reducing clutter.

- The existing Internet web services presented can be used to collect information for the purposes of OSINT and modeled to fit any format DHS desires. Their implementation tools need to be well thought out before being used, or else they will be useless.
• The exploitation of existing Internet web services will greatly benefit DHS in fighting terrorism and criminal activity in the US; allowing for better access to freedom (information sharing) and accountability from DHS.

Understanding the power of content can be very powerful and useful in the hands of capable IT specialists, language experts and analysts working at DHS. The exploitation of these Internet web services will bring about new ways to gather and analyze OSINT information. Those individuals that are in the technology department of DHS could very well implement and query these web services, with search engines like Google, Bing, Metacrawler, Dogpile, Yahoo, and so on. As I have described thus far, technology savvy individuals in DHS can program interfaces that can provide this concept or module interface and even work alongside private enterprises like Google to help them build a system where web content can be queried constantly. David A Umphress discusses this point:

For the technologically sophisticated audience, Google makes many of its services accessible programmatically through “application program interfaces.” In other words, a user can write software that draws upon Google's search features rather than having to use its Web interface. Thus, a user can have a highly specialized program for finding information based on an in-depth analysis of the results of Google searches, possibly combining Google results with those from other sources.65

---

CHAPTER 5

CONCLUSION

OSINT is indeed intelligence in the sense of providing important information for decision making on the basis of open-source insight. It provides information about our adversaries that they might not want us to have, but which they freely put out to the world to advocate for their views or actions. Properly interpreted, OSINT is likely to be just as enlightening as a well-informed secret agent or an image from an unmanned aircraft zooming in on a terrorist. In the end, what matters most in intelligence is the system’s ability to deliver sound judgments to decision makers. If OSINT aids in that process, then it is likely worth the cost and effort to collect and analyze it.66 Because much of the open-source information is available via the web, social networking sites, and the myriad of information sources that use the Internet as their delivery vehicle, the cost of this data is really not large, but the difficulty of gathering it is enormous in the context of general intelligence-gathering tools that historically rely on structured, intelligence-community databases.

As we have witnessed, the Internet has proliferated far beyond the expectations of DHS and the Intelligence Community (IC) and many of its domestic partners. DHS and many of its domestic partners will have to readjust their sights on this proliferation of the Internet in order to access freely available data more resourcefully and utilize these assets as they should be for intelligence insight. It would seem wise (advisable) for the Department of Homeland Security to institutionalize an operational structure to enhance how OSINT works for them, especially through the use of web services and cloud computing. This approach functionally utilizes the resources of others (Other People’s Money, OPM) such as Twitter, YouTube, Facebook, Google, Yahoo, and dozens of other Internet companies and assets, rather than DHS assets and resources, for most of the heavy lifting of data collection and management. When lives are at stake, DHS has only a very small window of opportunity to

succeed or fail; these very cost-effective tools can also be extremely rapid, enabling DHS to succeed and use knowledge as power on behalf of the people of the US and its government.

The following examples describe some representative ways that organizations (public or private) can utilize Open-Source intelligence more resourcefully. Actually, many groups in the corporate world do this every day, using with web services to prevent credit-card fraud with companies such as Fair Isaac\textsuperscript{67} that monitor most credit cards for fraud detection. Their use of near real-time information, cloud computing, and alerting is a wonderful example of using billions of pieces of information a day to provide actionable intelligence (fraud warning) that benefits corporations and credit card holders. Having DHS and the IC use similar workflows and web services could substantially increase their abilities to respond to terrorist threats, natural disasters, and other responsibilities with which they are charged by law.

The largest provider of web services in the world is Google, which provides Google searches across nearly the entire world. Their search engine provides searches on web information, but also images, videos, maps, news, shopping, books, finance, scholarly material, blogs, YouTube, and many other services. These requests go out to the cloud—anywhere Google assets can deliver the results—and return the results to the requestor. Queries can be done in many complex ways and can automatically bring in massive amounts of information to be aggregated, searched, and saved. Similar web services from other companies such as Microsoft, Yahoo, Amazon, and many others provide an extraordinary wealth of data that is available for query, though likely most DHS groups only use very simple queries and don’t develop databases based on these queries particularly in near real-time analysis of critical situations. One of the most robust of the Google tools is Google Alert, where Google reports to the user whenever a term or name or other search item is added to the web. Simply setting up a Google Alert, especially in multiple languages, is an extremely powerful way of gathering Open-Source data in near real time. This is simply using the Google infrastructure in an appropriate way to gather and report data of interest.

Facebook is rapidly becoming an extraordinary repository of information, particularly information showing relationships to others. Facebook is built on a recognition of human

\textsuperscript{67} For more information, see FICO, http://www.fico.com.
interactions, so the individual Facebook page functionally creates information about the people or organizations that are connected (Friends) with the “owner” of the page. This is an aggregate of the relationships that have been previously made and either agreed to by others (private Facebook pages) or with open data. Simply constructing a Facebook page and having the Facebook infrastructure gather and report the data, in a temporal and interactive order, can be deeply insightful. Many additional tools mine and analyze Facebook information for user accounts, so that extraordinary insight can be drawn with free or nearly free tools. HootSuite\(^ {68}\) analytics is one such tool that is a remarkable way to analyze this data and its relationship to other data such as from Twitter (HootSuite competitors: TweetDeck\(^ {69}\) and Seesmic\(^ {70}\)). As an example of HootSuite’s capability, HootSuite compile an Infographic\(^ {71}\) document based on consumer usage during Egypt’s and Iraq’s demonstrations talk about earlier in the introduction of this thesis (see Appendix).

YouTube is a favorite site for sharing massive amounts of information, including many videos of extreme interest to DHS and law enforcement. As an example, many terrorist groups use YouTube as a propaganda tool to advocate their positions, as do many criminal groups such as Mexican drug cartels. Cartel videos of forced confessions, torture, murder, and other criminal activities are a means of trying to strike terror into the lives of the general population and rival gangs or cartels. These videos are often extremely graphic and disturbing, but can provide an extraordinary amount of information only if they are found and utilized. Because YouTube uses text for searching, it becomes extremely important to use search terms that will bring up the required content. This includes search terms in the appropriate languages, rather than simply using English. Because many of the graphic videos are removed by YouTube, finding them before they are removed, or even coordinating with YouTube to view “removed” videos could open remarkable OSINT doors for DHS. The details of the videos, the number of people who watched them, the comments, and the linkages to other websites are also major sources of important information.

---

68 For more information, see HootSuite, http://hootsuite.com/.

69 For more information, see TweetDeck, http://www.tweetdeck.com/.

70 For more information, see Seesmic, http://seesmic.com/.

71 For more information, see Infographic Egypt, http://blog.hootsuite.com/egypt/.
Similarly, songs such as the *narcocorrido* “drug ballads” are another example of the enormous amount of Open-Source information that is available. Dates, people, locations, and lyric content are all part of this remarkable commentary on drug violence and activities. DHS analysis of such Open-Source information, which is nearly all in Spanish and also includes literally millions of comments about the songs, would be an extraordinary assistance to DHS activities. This would be especially true if the information by different cartels and their advocates within the entertainment world was contextualized in time, region, and activities. Tracking of the songs within the US would also provide insight into the cartel affiliation, such as by analysis of songs on the computers, phones, iPods, and other digital devices in the possession of people detained by DHS.

Twitter is perhaps one of the most remarkable Open-Source tools of usefulness to DHS because it is used so widely in the world, and it is being used for tactical actions such as “digital activism” and anarchy. Because Twitter can provide incredible situation awareness and command and control coordination of large groups of people, it has proved remarkably powerful in the recent efforts to overthrow the governments of countries like Egypt, Tunisia, Libya, Yemen, Jordan, Saudi Arabia, and others. Many of these Twitter accounts are open to anyone who wants to join, rather than being restricted to known sympathizers with a particular cause. However, if no one joins the accounts, they function as private command centers and control communications that can be used to the benefit of terrorists, anarchists, and others using these tools for their efforts.

Overall, Open-Source information can be and is transmitted via thousands of different services across hundreds of languages, and in forms varying from text, video, images, and songs. Gathering these data sets together, aggregating them into meaningful, structured databases, and using them for the benefit and protection of the nation is a massive undertaking. Such an undertaking requires immense computing resources, intelligence insight, and near real-time alerting and response capabilities. Because of the traditional government acquisition rules, it seems highly unlikely that government agencies such as DHS are currently able to successfully utilize this massive amount of information that is available, but hidden because of the enormity of it. Hiding it in other languages such as Arabic or Pashtun or even Spanish almost certainly keeps most of this information outside of the insight of DHS or other government agencies working to protect the US. This represents
an extraordinary opportunity for the nation and for DHS, but only if DHS chooses to embark on the utilization of this treasure trove of information.
REFERENCES


Department of the Army. *Open Source Intelligence (Field Manual Interim No. 2-22.9)*. Washington, DC: Department of the Army, December 5, 2006.


APPENDIX

SOCIAL MEDIA AND UNREST IN #EGYPT
and Unrest in #Egypt

**SAMPLE TWEETS**

**TRENDING HASHTAG VOLUME**
Worldwide

**EGYPTIAN MESSAGE VOLUME**
via HootSuite

**EGYPTIAN**
User Registration

---

Jan 25: Twitter blocked in the afternoon as demonstrations gain momentum. Curfews are broken by protesters.

Jan 27: Facebook also blocked. Egyptians find alternate access through tools like HootSuite and remote proxy servers.

Jan 29: The outside world helps organize resources for Egyptians to get their messages out via landlines, satellite, ad hoc ISPs and #speak2tweet.

---

Data: HootSuite & Ow.ly