The Report committee for Colter Roy Starr

Certifies that this is the approved version of the following report:

Considerations for Open Source Intelligence through the Lens of Information and Communication Technology

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: _____

Lynn Westbrook

_____

Lance Hayden

Considerations for Open Source Intelligence through the Lens of Information and Communication Technology

by

Colter Roy Starr B.S.

Report

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Information Studies

The University of Texas at Austin

May 2013

Considerations for Open Source Intelligence through the Lens of Information and Communication Technology

by

Colter Roy Starr M.S.I.S

The University of Texas at Austin 2013

SUPERVISOR: Lynn Westbrook

Open source intelligence (OSINT) has always been strongly tied to the information and communication technology (ICT) of the day. This paper is an examination of the current state of OSINT as it relates to ICTs by looking at overarching problems that exist across multiple types of collection methods, as well as looking at specific cases where there are issues, such as China and the Middle East, and ending with some minor recommendations on how to fix or minimize the issues highlighted.

Table of Contents

Introduction

The subject of this paper is open source intelligence within the context of information and communication. Examination of current methodologies and recent policy changes identifies professional, structural, and ethical implications inherent in using and adapting to technological progress. From here, the goal is to see where information professionals fit in this process and the intelligence community as a whole. The intelligence community is defined for our purposes as any of a number of groups whose goal it is to collect information, process it, and present it to policy makers in order to inform their decisions. Next, what is open source intelligence? The official government definition of intelligence is as follows.

> The product resulting from the collection, processing, integration, analysis, evaluation, and interpretation of available information concerning foreign countries or areas. 2. Information and knowledge about an adversary obtained through observation, investigation, analysis, or understanding. (U.S. Joint Chiefs of Staff, 2000)

As for the open source portion, this is not to be confused with the open source movement in computer software. Open source intelligence (OSINT) is collecting and analyzing anything that comes from publicly available documents or broadcasts. This covers newspapers, government documents, social media and broadcast media such as radio and television. OSINT is more related to collection than it is analysis. It can be thought of as collection without covert action (Davies, 2002). In the 1970s however, OSINT was

viewed as most useful when combined with other sources to create an all-encompassing form of intelligence because, in this role, it served to back up existing ideas put forward by other forms of intelligence (Davies, 2002). Information and communication technology (ICT) is defined here as the devices that allow for a person or a group of people to send a message to another person or group of people.

In the past, open source intelligence came in two main forms, print and broadcast media. The first was print including newspapers, government documents, and straight out propaganda documents. In many areas of the world they are still the only sources available. In cases like that of North Korea, this is by choice. Other times countries and people of a given area lack the technological means to produce much if any open source data in an online electronic format due to a lack of financing and infrastructure. A major issue with this type of intelligence in the past was that it mainly provided information after the fact. For example, if someone was trying to prevent a bombing, a newspaper article reporting on the aftermath of that bombing was not going to be useful in preventing the event. One of the goals of intelligence is prediction. Since OSINT did not, on the surface, appear to be as useful at predicting events it was seen as a lesser pursuit.

Open source intelligence focused more on giving the culture and governmental undercurrent of a particular area. If an intelligence officer wanted to get a general feel for what issues were important to another secluded country, propaganda was a good source for that type of intelligence. North Korea is a good current example of this idea. Official state documents and broadcasts from North Korea are often not viewed as accurate at face value. Instead, they give other information such as what is currently being viewed

as an important topic by the government. If they are talking about having a plentiful supply of food for example, it may not be that they actually do but it does mean that food supply is important to their current thinking (Mercado, 2003).

During World War II radio broadcasting became a widely used form of open source intelligence. Radio broadcasting is not to be confused with military radio transmissions which fall under a different type of intelligence; the open source form is what the common person would listen to for news and entertainment such as President Roosevelt's fireside chats. This area was monitored by the Foreign Broadcasting Information Service (FBIS), formerly the Foreign Broadcasting Intelligence Service (Mercado, 2003). The change from intelligence to information both indicates an increase in breadth and a decrease in credibility. Where intelligence was often viewed as a polished product, information was seen as raw data.

Throughout the Cold War print and broadcast monitoring continued to dominate. Television was added to radio in the broadcast realm. Open source officers tried to prove the usefulness of their highly accurate product, but other areas of the intelligence community were still perceived to have greater importance. Mercado (Mercado 2003) points out an example of this phenomenon. "FBIS [Foreign Broadcast Information Service] and FDD [Foreign Document Division] officers began discerning signs of the Sino-Soviet split from their readings of propaganda material in the early 1950s." Yet other sources, such as human intelligence or HUMINT, were taken more seriously even though they ended up with incorrect predictions more often. The result was one of the

early intelligence failures for the United States to predict the actions and status of communist countries. Open source intelligence remained the black sheep of the intelligence world until two things took place: the invention and commercialization of the internet and the September 11, 2001 attacks.

The internet and 9/11 caused the intelligence community, and those relying on the intelligence community to inform their decisions, to take a closer look at open source data. Why wasn't it being used more thoroughly and often? Mercado (Mercado 2003) sums up the surface argument for the use of open source intelligence. "One can gather more open intelligence with greater ease and less cost than ever before." The problem is that ease and low cost come with their own sets of problems such as information overload and ethical concerns like privatization.

While steps have been taken to make better use of the wealth of knowledge that is available to all, there is still resistance from the current system both ideologically and technically. "Too many people still reject OSINT as intelligence. Worse, too few are able to gather and exploit open sources. (Mercado, 2003)" That said, change is happening with regards to its use in the intelligence community. Open sources are now being taken more seriously because of new and developing technology, specifically the internet. While this may seem like something self-evident, the uses of this new medium were slow in coming and are slower still in implementation. The internet has the ability to be a treasure trove of information that can be mined, analyzed, and shaped into useful intelligence products if properly handled in terms of both of policy and technology. No longer can or should

open source intelligence be viewed as an after-the-fact distribution of data on an event from a certain point of view. The internet allows publicly available data to be gathered before events occur to help make predictions just as well, if not better in some cases, as more secretive traditional types of intelligence gathering. Blogs, for example, while possibly falling into the pit of bias, are able to give real time, on the ground coverage of events that may or may not be reported by the news or picked up by HUMINT. Blogs, and most other internet based social media, allow analysts to collect information on the feelings and beliefs of the people in a particular region. Facebook, Twitter, YouTube, forums, message boards, and other information media provide a wealth of open source data. The list of possibilities for open source intelligence gathering on the internet is long and dynamic. While information can be created online in an instant it can just as quickly disappear.

Information on the internet is fluid and temporary. It has a short half-life. If, for example, a blogger posts a piece and later decided she is unhappy with it she can remove it. Unless someone made a copy of that article in the time it was up, it is no longer available to the public. There are also space restrictions for online information. If that same blogger works for a major newspaper, the post may appear on the front page of their site for the first few days after it has been uploaded. After that initial time period it can be moved to a different location, be that the blogger's personal space or storage, because it is less relevant to the current news cycle and front page space is needed. The article still exists at this point, but finding it becomes difficult. The path is changed and possibly broken. YouTube presents legal reasons for the disappearance of data. If a video

receives a complaint of copyright infringement it is removed whether or not this is actually the case. YouTube has an appeals process, but it is used to get information restored not to prevent it from being removed. The internet is not a place of permanence.

The global positioning system is a tool used throughout the intelligence community. . Most people know it simply by its acronym GPS. GPS data is increasingly becoming publicly available. It is a feature of many social media sites by allowing the user to make updates from mobile devices detailing their location. In many cases these updates tag the GPS data of the user at the time of the post. Depending on the site, the user may not even be aware this is happening. The system may be automatically sending data unless the user chooses to opt out. Foursquare is a social media site that is built entirely on GPS data. Using this open source feature, tracking the location of an individual can be done through an internet browser without needing to go through the legal system to obtain a court order.

The traditional low-tech open source strategies of print and broadcast monitoring are effective when applied carefully to the nature of local information. For example, North Korea uses print and broadcast media heavily due to their systematically censored and limited amount of information. On the other end of the continuum the United States uses the World Wide Web media heavily due to their deep internet penetration and commercialization of data dissemination. Even in industrially developed countries however print and broadcast mediums' use generates meaningful data. For example, in the U.S., a 2012 Pew survey notes that 42% of the 65 and older population, a group with

significant political input, still do not use the internet (Pew Research Center, 2012). Low tech monitoring will have a place in OSINT for some time to come.

OSINT within the Broader ICT Framework

Why They Matter in the Intelligence Community

During World War I and II, when the intelligence community in the United States was still getting its footing, there was a clear target on which to collect intelligence. Germany and its supporters were the focus. There was a name, a face, and borders. It was clear on whom to gather information. During the Cold War there was again a clear target with a name, face, and borders, and again collection of intelligence had a focus. The Soviet Union and other communist states were viewed as a threat about which intelligence products were needed in order to make informed policy decisions

In the current climate however, the entities concerning policy makers do not come in the form of nations. They come in the form of cells and networks. There are no borders. There is no face. Sometimes even the name is unclear. The formations concerning policy makers are also tech savvy.

> The security landscape was rapidly changing due to the amorphous nature of unconventional non-state threats such as al-Qaeda and the Taliban, along with their increasing ability to use technology for nefarious means (Chomik, 2011).

Technology is no longer cost prohibitive. Take flash drives as an example. Ten years ago a 256mb thumb drive cost roughly $150. A device capable of holding 44 times that much can be purchased for about $15 today. Those viewed as threats are now built like the

internet meaning that they are made without a central location so that if one node goes down the system can reroute and continue to function. Like the internet, they are made to withstand attack. An arms race in the way that it occurred during the Cold War with the Soviet Union now makes little sense. Open source intelligence allows for the collection of data in a spread network style more akin to the way in which current organizations operate. More likely than an arms race is a race in terms of technology as that is more effective at combating and gathering information on decentralized and technologically advanced groups. This does not refer to a physical hardware race either, but instead one of software (Chomik, 2011). It is easier to infiltrate a network electronically than it is physically. Risk of things like double agents and loss of life can be reduced.

All of this matters because the availability of information is growing and changing. It does not live in the same place it did in the past. It is important to understand and adapt to this idea. Failure to do so will result in misinformed policy and mis-allocation of resources. This will lead to failures such as missing an attack or losing out on the opportunity to help those in need.

Essential OSINT Design Factors

Many of the issues facing ICT's and OSINT deal directly with the information portion of the information and communication technology. Two prime examples of this are transparency and information overload. Transparency focuses on how visible the collection of intelligence, and the means of which it is collected, are to the parties involved. Information overload deals with the volume of data available to agencies and how its volume affects the quality and accuracy of intelligence products.

Finding the Appropriate Level of Functional Transparency in the OSINT Data Collection Process

Transparency in intelligence gathering is a double edged sword. It is one that is difficult to decide which edge should be the most sharp. OSINT is as clear an example of this idea as it gets through its use of public data. While the internet can be used to find out more about a threat, the threat is able to use the same sources to find information about those they threaten. It is a warped version of that old Friedrich Nietzsche quote. If you stare into the internet long enough, the internet stares back at you. In the intelligence community, transparency is a bi-product of the internet instead of a goal. "Transparency does not mean that everything is completely open, nor that it should be (O'Connell, 2004)." The idea that it is unintentional raises the question of how much information should be allowed to be open source. It is a question that speaks to the very nature of information by looking closely at the purpose of its creators. The current system in the

United States is somewhere between transparent and opaque. The more open source intelligence is available and is used, the more transparent things become. It can be a frightening concept to those who have had a system in place for years that has intentionally been kept as opaque as possible. This explains why many within the intelligence community, and the government in general, hesitate to use some of the tools the internet affords.

> Neither the Intelligence Reform and Terrorism Prevention Act of 2004 nor the 9/11 Commission Report devoted much time and attention to science and technology and their roles in U.S. intelligence (O'Connell, 2004).

Add to this statement the changing attitude of the public in the direction of openness in information, and the scale tips further is favor of transparency. Other INTs are also feeling pressure to become more open. The likelihood that it will be successful in other areas is not nearly as high, however. This is due to the fact that they do not need to be transparent to function. Higher opacity is necessary for other forms of intelligence, specifically HUMINT, to function at all. Open source intelligence depends on a certain degree of transparency, so that is where change will be seen first.

Fear of transparency is understandable to a point. Whether or not intelligence should be transparent is not the issue here. The concern is how people have gone about making it transparent. One matter of contention is Wikileaks. This is not a condemnation of Wikileaks. It is not a question of whether Wikileaks is a champion of justice or evil incarnate. What matters is how Wikileaks went about making public the data it had in its

possession. It had a chilling effect on how intelligence agencies went about using ICTs and on the level of transparency that they were willing to put in place. By pushing the information out in the fashion it did, Wikileaks functioned like a parent teaching their child to swim by throwing them in the deep end of the pool. Sometimes that child will learn to swim. More often than not, that child will develop a fear of water. The intelligence community did not learn to swim (Schroeder, 2011). This fits with the concept held by many in the post-9/11 world. "In times of national or social threat, history has demonstrated that governments often expand surveillance and other powers at the expense of citizen rights (Strickland, Baldwin, & Justsen, 2006)." In addition to Wikileaks, which is arguable as a threat, there are some very real dangers present to those producing open source intelligence. For example, in 2011 alone four bloggers were murdered for putting out content dealing with the drug war in Mexico (Kessler, 2011).

Finding Functionality in OSINT Staff's Ability to Manage Data Processing Information Overload

Information overload is when so much data is coming in or is available for use that it becomes difficult and overwhelming to organize and search for value. For decades people have struggled with deciding how to deal with all the information available to them. Along came the internet and that problem became exponentially worse. At least part of the problem is training. Intelligence officers often lack the training for dealing with large amounts of raw data.

> The first challenge is that U.S. intelligence has been and remains overwhelmingly collection-centric, with insufficient attention paid to the creation of end-to-end (or sensor-to-analyst) architectures that will be needed to create useful and actionable intelligence (O'Connell, 2004).

With the information provided by the internet available for use, the question of what to do with it becomes central. People and systems with the ability to sift through the data and pull out the portions with value for intelligence purposes are needed.

Through the lens of information overload, capitalist systems make for harder targets. There is more competition in websites, broadcast stations, blogs, etc. A country like North Korea makes an easier target since there is only a handful of OSINT to be had. Stephen Mercado makes the idea of gathering from a small pool sound like a positive point; however, this is only a positive if there is a limited number untrained staff available (Mercado, 2003). If there is a staff of individuals in place who are well trained in dealing with information overload, the intelligence that can be gained from the internet is sizable. However, Mercado's point is not invalid. There is a point at which information overload is a more effective way to hide data than actually censoring it and keeping it secret. Eventually the volume of information will reach a level where this will be the case; however, currently and in the short-term future it is not.

Information overload is used as a counter to open source intelligence. Technology has not yet reached the point where software is able to comb the entirety of the internet and catch every piece of data that might be of use in creating an intelligence product.

Going back to the examples of North and South Korea for a moment, while South Korea has a larger amount of open source data available, it would be easier for information to slip by an agent trying to collect intelligence than it would if that same agent was looking at North Korea. In this scenario North Korea is a known unknown. When collecting intelligence from open sources, the collector can be aware that it is heavily edited and censored. The officer is aware of what it is they do not know. South Korea, along with the United States, is an unknown unknown to a certain extent. There is so much information being produced that it becomes hard to tell what is being missed (Mercado, 2003). The idea of known unknowns and unknown unknowns has been around for a long time, but it was popularized again in recent years by Donald Rumsfeld. From there it began to be used as a joke. It is however a valid philosophical idea.

There is a split occurring within information overload between human overload and technological overload. In one direction is the traditional view which is an overwhelming amount of data regardless of format. It is an human issue. The information is available, but the user can not find what they are looking for within irrelevant data. The second area is technological information overload. In this situation, the problem is not finding specific information within a larger set, but instead it is problem processing available data through a system. An example of technological overload is the Chinese social media issue. There is screening program in place to make sure a post does not break Chinese law. If it is overwhelmed by the number and size of the posts coming in, this second form of information overload will occur (Elgin & Einhorn, 2006.)

Information overload is an issue that is set in information, but affects technology. While it is a concern both online and off, the digital realm plays a large part in how it is handled. The next section focuses on the technological side of intelligence gathering and analysis and some of the major players and problems within it. By examining these areas the link between technology, information, and OSINT becomes clear.

The Technological Tools for Data Gathering, Data Processing, and Data Creation

There are various aspects of technology that effect OSINT from hardware to software. Social Media is the central open source intelligence tool. Another area that deserves mention is location data; however, this is a secondary focus. Location data illustrates where and how data is used and deals more with hardware. Each of the technological areas covers certain, if not all, of the ideas of data gathering, data processing, and data creation. Data gathering is the collection of information created and put out in a place that is considered open source. Data processing is the ability of the tools to take information created elsewhere and form it into something useful for intelligence purposes. The tool may be designed to do this with or without the user's intent. Data creation is the ability a tool gives its users to create raw information. Within the realm of software, the key form is social media. Location data focuses more on hardware. The capabilities of these products effect OSINT through their many to many interactions.

Social Media

Social media covers data gathering, processing, and creation. Data is created by users, it is gathered in one place, and it is aggregated into groupings and categories. The focus here is on social media that is popular in the United States. For social networking the main example is Facebook. There are sites similar to Facebook that are more region

specific. In Russia there is V Kontakte. In Eastern Europe they have Odnoklassniki. China has its own social network called QZone (Cosenza, 2012). Twitter, a micro blogging site, is a beast all its own. In addition to Facebook and Twitter there are blogs and forums. Blogger, wordpress, and tumblr are just a few of the more popular blogging sites.

Can users of social media expect some level of privacy? Facebook, for example, has privacy controls. If a user adjusts them properly, they should be able to expect that certain, if not all, of the information they put up on the site will not be open to the public. On Twitter there is a very blunt setting to handle privacy. Users can protect their tweets or they can let them be. While Facebook still retains its usefulness in private mode, the value of Twitter decreases. The point of Twitter is to put something out there for others to read. If users are using Twitter as intended, which will be defined as using it with the default settings, can they expect any level of privacy? The answer on the surface is no. An unprotected Twitter account is the same as someone talking loudly on the street. Passersby will hear what is said. An unprotected account does not prevent people from thinking they should still have privacy. For the purposes of open source intelligence gathering, they potentially have a point. Intelligence gathering would then be akin to following someone around on the street with a tape recorder. It still might not be illegal, but the complaint of encroaching on privacy would be legitimatized.

With the example of Facebook specifically, can the data be considered publicly available? Even with privacy settings turned all the way up, it is still unclear who actually

owns the data users upload to the site. It is not possible for a user to delete a Facebook account. Accounts can be deactivated, but all users need to do to get back on is log back in. All the information originally entered is still there. If users can not get their data back, ownership has shifted to the company. At any time Facebook could sell, or give away, its user data to the highest bidder (Protalinski & Blue, 2012). If so that data would be useful intelligence. Is this open source? A newspaper is open source, but anyone wishing to read it will still have to pay for it so in that sense it is. Again to clarify this example of data control is as things are in the United States. Countries like Germany for example have different laws regarding the rights of citizens as they relate to corporations. Facebook has a high degree of global saturation which is both potentially positive and negative for the purposes of OSINT. On one hand a large reach creates a one stop shop. It is easier for intelligence agencies to create tools and algorithms for pulling out valuable data if the focus is on a single platform. On the other hand saturation on a global level creates possible problems. Facebook could decide to prevent the use of its data. In this way Facebook is unlike traditional public media. In the past once a newspaper is printed or a broadcast sent, it was available to everyone even if the creator decides after the fact that they wish it were not. In the current phase of technology the ability exists to pull back information. While Facebook is not the creator of the data, they do manage it and have control over its presentation once given to them. An agreement would need to be reached between collectors and Facebook. If such a deal were reached international laws would need to be considered. As of right now, there is a document proposing changes to Facebook's data use policy. Within that document, there is a portion that states users can

choose whether or not to make a post private or public, but that it will not always be present and when it is not the post in question defaults to public. It is unclear what classifies a post as forced public or choice public. "Your settings do not control whether people can find you or a link to your timeline when they search for content they have permission to see, like a photo or other story you've been tagged in (Facebook, 2012)." Based on statements like this one, there are ways around privacy settings. Either way company policy and foreign law regarding social media sites is important for open source intelligence collectors so they know what data they can use and what they cannot as well as how to use it if they are allowed access.

Facebook allows for many to many public and many to many private interactions. Mass private messages can be sent out just as they can in e-mail. This is a feature that differentiates Facebook from other forms of social media. Twitter allows many to many interactions to occur publically, but not privately. Protecting tweets is still public because anyone a user has allowed to follow them can see what that user is saying to anyone else. The privacy setting allows for restricted public, but the data is still public. Forums and blogs are much the same way. Most social media does not allow for many to many private interactions to occur. Facebook and e-mail are the two main players that effectively are able to handle this type of communication, and e-mail is neither social media nor open source.

In the historical discussion earlier, the two areas where open source intelligence officers focused their efforts were print and broadcasting. Print and broadcast are

communication, but they are not interaction. A message is sent and a message is received. This is not the way ICTs work today. Today almost every type of communication technology out there sends messages in not just one, in many cases not even just two, but many directions.

> ICTs provide two-way channels for those who seek or receive information and also permit many-to-many communications. By incorporating feedback, political activity becomes more personalized and enhances the potential for engagement (Robbin, Courtright, & Davis, 2005).

Like many of the areas of open source intelligence, sending messages to multiple recipients is a double edged sword. If more people are engaged in a conversation that can be publicly viewed, then there is more information for collectors to go through. Technologically there are numerous ways for information to travel many to many with feedback at this point in time. E-mail, blog entries, forum posts, and social media are just a few examples of places where someone could communicate to a large group of people and get feedback. Communicating to large groups also means there is a higher chance that something will be missed due to information overload. Data traveling in multiple directions has its advantages. In an environment where broadcasting is the key form of dissemination, there is only one source to follow. If communication is occurring in other directions, it can be followed to other sources. Multiple data sources leads to unexpected avenues of collection.

Many to many transmission begins to bring up the ethical and responsibility concerns involved with OSINT. Large numbers of messages can be sent and received by large numbers of people. Who has the right to collect these messages? Can the users expect any level of privacy? These are some of the considerations that will be discussed in the following section which centers on social justice.

Blogs are the online source that most clearly exemplifies the new face of open source intelligence. Blogs are the newspapers of the internet. They range from professional journalistic reports to the opinions of a single individual. Blogs can be used in the traditional sense similar to a newspaper. In fact, most major newspapers have their own blog at this point. Within professional blogs is information on what events are happening and how those events are being reported. On the other end of the spectrum there is the personal blog. Personal blogs could be anything from the thoughts and feelings of a farmer in a war torn region, a tween in Middle America talking about fashion, or a high level executive in a foreign country. Just from these examples, it is clear that some blogs are going to be more useful than others in compiling an intelligence report. By analyzing the personal blog, a general idea of the feelings of the common person can be obtained. It is also possible to see what the major news sources in an area are not reporting. Blogs are clearly open. Going back to the person on the street analogy, a blog is the equivalent to someone giving a speech to a crowd. Blogs are so open in fact, that their transparency can be dangerous. Multiple bloggers have been killed for writing about the Mexican drug war since 2011 (Kessler, 2011).

Blogs are also one of the major sources of information overload. In places like China this is becoming a problem. China's system of law does not prevent the creation of blogs, but it does require posts to be screened before officially being posted online. Since anyone can have their own personal blog or even video blog, the amount of data requiring review is large. Video blogs are becoming a concern for China's screening process (Elgin & Einhorn, 2006). Video files are larger and thus take longer to process from a technological standpoint. There is also more to analyze for content in video format than in text. Chinese law prohibits certain content. In text it can only be in the words. In video it could be in the words, the background, the physical actions that take place in the video, or a number of other things. While video makes things more difficult for China to screen, it also makes things more difficult for open source collectors and analysts to catch.

Location Data

Many social media and mobile based applications now have the ability to add location data to a post. This location data covers data creation and gathering. Through a user's IP address or GPS coordinates it is possible to get a close approximation of where they were at the time of the entry. Recently, a study was conducted on the behavior of those using such social media sites to see how much they were willing to share about their location. The study found that, "location information is preferably shared on a need to know basis, not broadcast (Wagner et al., 2010)." However, they go on to say that, "disclosing location at the granularity of city is perceived as disclosing nothing (Wagner

et al., 2010).” This shows a disconnect between what is now thought of as sharing and what was previously thought. Location information is important intelligence. Knowing just the city someone is located in can be enough. If disclosing the city is seen as disclosing nothing then the first finding that location is shared on a need to know basis is not as important as it originally seems because disclosing the city is not seen as disclosing the location. Location data also allows for a rudimentary census. Using the information from a number of individuals it is possible to estimate the number of people in a given place at a given time.

Another finding dealt with how willing users are to share the location of others. It turns out they are not very willing. “Users are more cautious when sharing others’ location (Wagner et al., 2010).” People are more likely to tell someone where they are than they are to tell where someone else is located. The way the study conducted this portion used an extreme example, however. The researchers posed a question about giving the location of someone who was potentially having an affair. The finding was that people would err on the side of protecting the person whose presence was in question. In a more everyday case, it would be useful to see if the finding that people would withhold the location of another person still held true.

One reason certain participants in the study were reluctant to give up their location was that, “locations are associated with actions (Wagner et al., 2010).” This means if someone shows up as at the office on a social media site, people assume they are working. If people show up at a restaurant then they are eating dinner. If they are at home

then they are not busy. The point was made that showing up at home if you are working from home that day is a negative and that if they could say working from home instead of just at home then they would be more inclined to share that type of data (Wagner et al., 2010). This is another bit of knowledge that OSINT agents, particularly analysts, should not overlook. Just because a person is at a certain place it does not mean that they are engaging in a prescribed activity.

Social Impact and Responsibility of OSINT in its use of ICT

When looking at open source data on the internet, it is easy to get mired down in the specifics of what can be done while overlooking what should be done. Intelligence collection is a broad area with many ethical shades of grey. By looking at the current system, it is apparent that there is a need for further examination in two main areas of social impact. This section focuses on the idea of social impact by looking at the responsibility and ethical concerns involved with creating and collecting information for intelligence products.

Responsibility for Information Creation and Use

If dangerous information is put up online who is responsible for it? For that matter, who decides what information is dangerous? Say for example someone in Texas posts plans for a violent coup by Mexican drug cartels of the Mexican government. Who is responsible for this data? Is it the person who posted it? Is it the forum owner? Maybe it is the cartel's responsibility. An intelligence officer who comes across this data must take these responsibilities into consideration when deciding how to handle this situation. If this is a crackpot attempting to incite violence the agent may act much differently than if this is an agent of one of the cartels putting out a call to arms. Online it can be hard to tell what the context is behind information. The context of the source is ambiguous. Other areas deal with the problem of uncertainty as well, take HUMINT for example, but none as often or as heavily as OSINT. Is anyone truly responsible for information? This essentially gets back to the "information wants to be free," idea that often rises from

technology activists. Protection of data is another area of responsibility that is important when talking about ICTs. To what degree is the creator of a site responsible for protecting the data of its users?

> The technological community needs to answer the criticism of science and technology studies, and must react to voices that highlight the ideological attractiveness of various alternatives to what is labeled a "western model" of science and technology (Resnyansky, 2009).

Ethical Concerns of Collecting and Using Information

Each of the topics previously under discussion raises its own ethical problems. There is a delicate balance between the level of transparency that is deserved and that which is actually possible. By making information publicly available, there is a fear that whoever is giving out the data is putting his or herself at risk. The risk taker could be an individual, a corporation, a government, etc. Information still needs to be made available, however. When compiling an intelligence report for about a company someone is considering investing money in, that investor has a right to know certain things, like if the CEO is under indictment on embezzlement charges, in order to make an informed decision. People have the right to know what their elected officials are doing in their name. There is an amount of transparency that is necessary for a democratic society to function properly. There is also a time for a more opaque approach. Once a country has

engaged in war, whether a citizen agrees with the war itself or not, making troop movements and tactics public would only serve to make a dangerous situation more dangerous. The internet would make it easy for the military to keep citizens up to date on every move in real time. That does not mean that they should. The line is unclear. Wikileaks lands on the far end of the transparency spectrum. Wikileaks was and is the epitome of a loss of control (Schroeder, 2011). It is easier for governments and intelligence agencies to feel like control is in place when things are opaque. Just because Wikileaks believes in the ultimate level of transparency, does not mean that fear of it should cause intelligence agencies to go to the far opaque side.

Transparency does not automatically fix any issues that arise in the intelligence community nor does it instantly make them more ethical. This concept is made clear in the system of überveillance. Überveillance is basically the idea of an ever present "big brother". There are cameras placed throughout cities. People are openly monitored electronically 24 hours a day, seven days a week. While the idea of überveillance itself is not new, at least as old as George Orwell's 1984, the plausibility of it is. With the advancements that have been made recently in technology, it is not impossible that a world where an individual could be openly identified and tracked at all times could exist. This has been proposed in countries other than the United States due to legal and geographical reasons. Geographically, the United States is a big place and monitoring it in its entirety with cameras would not be feasible. In its own way, überveillance is transparent system. It does not hide what it does. It is open (Michael & Michael, 2007).

On the issue of government versus the private sector, there has been talk of using private contractors to get around codes that prevent American intelligence agencies from collecting data on American citizens. In the online environment, borders are fluid. The ethical issues come in to play when deciding how to deal with this problem. Going to private contractors and simply circumventing the law is questionable at best. By using outside organizations to do the job government agencies cannot legal do is equivalent to that agency breaking the same law. The private sector has much less oversight. The general public has virtually no say in how private sector intelligence agencies conduct themselves. Oversight and accountability are important in all areas but few more so than the field of intelligence. A more ethical approach would be to reevaluate the existing policy to create a policy that works properly instead of using a temporary fix (Hosenball, 2008).

Cost polarization is another area with ethical problems. By allowing the best intelligence to be bought by those with the most money and leaving the rest to fend for themselves, a set group of moneyed individuals is kept in a position of power whether they deserve to be or not. Wealthier countries have always been able to afford better intelligence. The problem comes in putting the ability to control intelligence in the private sector thus giving more political power to those with less oversight and accountability. While increased power in the private sector may happen either way, allowing it to happen and helping it to happen are completely different (Hosenball, 2008).

Ethical issues with OSINT occur in the real world through the policy set by governments and organizations and its implementation. Understanding issues with these areas helps to clarify the world in which the ethical issues discussed above live. The next section gets into these areas both on a global and national level.

Policy and Implementation of ICT use in Open Source Collection

The way in which ICTs and OSINT come together hinges on the policy of the government or organization using them and how that policy is implemented. The section begins with international concerns such as the "Great Firewall of China." Next it moves on to the relationship between the government and the private sector both in terms of technology and intelligence production. The last portion looks at how the use of information professionals could improve the current system.

International Considerations when Collecting OSINT

A very specific issue for open source intelligence when it comes to communication technology is China. The phrase "Great Firewall of China" is ubiquitous on the internet. This term refers to the technological system China has set in place to regulate and monitor the flow and type of information coming into and originating from its borders (Elgin & Einhorn, 2006). Much of this data is of the type that would be useful to OSINT officers. Evidence that China prevents that sort of information from leaving the country came in 2010 when Google reported a cyber-attack which emanating from China attempting to pull information from the accounts of various Chinese human rights activists. A new sort of war has even been postulated in the form of cyber-warfare could go so far as to cause a new sort of cold war. In this cold war, however, the race will not focus on nuclear armament, but instead on technological superiority. (Hartnett, 2011) This idea underlines the importance of harnessing the internet for intelligence. China's official stance is that it does not intrude on political information. There is still ambiguity

around the topic of what is fair game, however. This uncertainty makes the job of the OSINT analyst difficult. It adds an additional layer to the criteria for trustworthiness in data. China has, in the past, done very well with the problem of information overload. It is theorized that there system will not be sustainable, however. There could be a time in the near future where China's system no longer functions due to the sheer magnitude of data that is pressing against "the Great Firewall (Elgin & Einhorn, 2006)."

China also is a clear example of the disconnect between internet use and internet penetration. There are more people online in China than in any other country in the world. China has more than double the number of users than the second highest country which is the United States. This gives the impression that it would be easier to construct a picture of China than of the U.S. from open source data, censorship aside. However, while China has roughly 538 million users and the U.S. only has about 245 million, the penetration of the internet has only reached 40% of the population in China where in the U.S. it is at 78% (Miniwatts Marketing Group, 2012a).

The Middle East is another area where open source intelligence is complicated with internet usage numbers. Iraq is still recovering from war. Before the Iraq war, there were guidelines in place as to how the government would allow the internet to be used. In 2000 only .1% of the population was online. That number saw a jump after the government fell. Iraq went from 12,500 users in 2000 to 325,000 in 2010. This is still only 1% of the population. The amount of OSINT that can be gathered through internet-based means is not going to be as helpful in creating reports on Iraq. Pre-war

governmental control mixed with post-war infrastructure damage has kept Iraq technologically behind. Traditional methods would be more efficient for collecting intelligence (Miniwatts Marketing Group, 2010a).

Iran has heavily censored access to the internet (Tait, 2006). There penetration numbers are similar to China with around 40% of the citizens having access. From the stand point of open source intelligence, this is a better number. It means there are over 33 million Iranians who can go online (Miniwatts Marketing Group, 2010b). The problem here is censorship. Collecting intelligence on a censored system raises the question of which is better, a few sources from a place with little censorship, or many sources from a place with heavy censorship? Mercado argues for both and points out that the two are both valuable because they provide different types of information (Mercado, 2003). Looking at what type of data is being blocked by a country's government is raw intelligence in itself.

Israel presents interesting data when viewing their internet usage through the lens of open source intelligence. Their internet penetration has actually decreased from 2008 to 2012. A smaller percentage of the population uses the internet now than did four years ago. In 2008, 74% of the population was able to go online. Now that number has fallen to 70% (Miniwatts Marketing Group, 2012b). In all surrounding countries, including those with violent civil unrest such as Egypt and Syria, the penetration of the internet is steadily on the rise. Unexpected information like this is the type of data that would be good to share with other forms of intelligence. With data that is unclear, the all-sourced

intelligence method mentioned earlier can be helpful. In this context, the data on decreased penetration is confusing, but looked at along with other data it might make more sense (Davies, 2002).

Government Intelligence interactions with the Corporate World

There is a disconnect between the government and the private sector when it comes to OSINT. The mismatch comes from discord on the subject of regulation. There are laws in place to keep governmental power in check when it comes to the use of technology to procure information. Because of the legal chasm that exists between the two, tools created by one may not be able to be used by the other. The new world of the internet is a corporate world. A corporate world creates corporate tools. In doing so the world of government intelligence is alienated. The system in place encourages use of contractors instead of government agents.

Information and communication technology has reached a point where the private sector has outpaced the intelligence community in its abilities. Cost polarization is what occurs when the private sector offers certain information that has been collected, curated, and polished into a finished report at a premium from data that is available in a raw form for free or little cost. Paying a contractor to collect and create finished intelligence has its benefits Chief among them are speed and ease. The problems that

arise from using contractors are cost and trust. Can the government really trust that the information is being presented in a factual non-biased fashion?

The free market drives and guides technological growth. What sells drives technology and what technology is available drives OSINT. Since the advent of the internet, communication has been the driving force. From the early days of forums and chat rooms to the modern tweets and wall posts, communicating in an entertaining and informative matter has driven the internet. None of these technologies are owned or regulated by governments the way that broadcasting was in the past. The difference in regulatory control has hampered the ability of government agencies to collect open source intelligence. There are different legal standards government agencies must abide by. In an article by Mark Hosenball, he discusses the use of private contractors to get around this block. OSINT is causing government intelligence agencies to reconsider their policies based on new technology.

> The charters of the C.I.A. and the defense intelligence agencies prohibited them from spying on American citizens, under the logic that the intrusive tactics needed to investigate foreign threats would violate constitutional rights if applied at home (Thompson, 2006).

As the quote states the C.I.A. is prevented from spying on American citizens. The internet has globalized communication in such a way that it is hard to think of a social media site that is not going to have Americans involved. If tweets are monitored, Americans are monitored. If intelligence agencies go to Facebook they are looking at

Americans. The point is not that Americans dominate these areas. The point is that social media sites have a big enough presence to make it nearly impossible for someone to collect open source data from them without, unintentionally or otherwise, "spying" for lack of a better word on Americans. It is no longer possible for organizations like the Central Intelligence Agency to operate under the same policies they have in the past. Even the post-9/11 retooling of the intelligence community has failed to fix this problem (O'Connell, 2004). The choices for intelligence agencies are to go about collecting open source intelligence in a new way or change the policy in place. In the end a combination of the two will likely be necessary.

The idea of government versus private and old versus new policy becomes even more important if a cyber-cold war were to happen (Hartnett, 2011). There has been a fundamental shift in technological development since the formation of government intelligence agencies. At one time technology was developed to meet the needs of the government more so than that of the consumer. This has flipped in the last 20 years. Consumerism is the current driving force for advancement. During the Cold War, a government focus on weapons technology made an arms race was possible. In the current environment of consumer focus on information and communication technology, a physical arms race will no longer be valid. "Technology is a powerful factor that can either support necessary changes and provide useful solutions, or amplify undesirable trends and practices (Resnyansky, 2009)." Right now the tools for the intelligence community to function under these conditions do not exist.

Open Source Intelligence and Information Professionals

Information professionals are in a place to take a formative and integral role in open source intelligence gathering. To begin, they must look at their responsibilities within each of the above areas. One of the key goals across the board is education, specifically education of policy makers. It is their duty to make sure that technology like the internet is understood as something more than a series of tubes. It is also important to help others in the intelligence community understand how to deal with information. Educating intelligence agencies on how to deal with information will go a long way in helping to deal with information overload. Education needs to occur at all levels from local police departments to foreign policy makers.

Another area that information professionals need to cover is monitoring. Monitoring deals more heavily with technology than any other area due to the quick evolution of product that it covers. Information professionals must keep track of technologies as they are created and evolve. In order to do their job properly, it is necessary to have an understanding of both what new communication methods do and how they do it. Information professionals must also be up to date on how already existing technologies change. A perfect example of technology evolving is the cell phone. It went from calls to calls and texts to a mobile internet device all in the span of roughly ten years. Information Professionals need to keep track of this change so that they can figure out what it means in the world of intelligence.

Information professionals have been dealing with information overload for years. It is their responsibility to aid those collecting vast amounts of data for intelligence with the knowledge information professionals have accrued on the subject. One of the major skills an information professional possesses is the ability to curate a collection of useful data for a target audience. This is exactly the sort of skill that is needed in the current intelligence community. Information professionals have the task of both working at solving the problem themselves and teaching others what they can do to mitigate the overflow. Most of what is out there floating around in the internet is garbage data. It is necessary for a curator to set up a system to go through and decide what is worth keeping and what can be thrown away. Information professionals allow governmental agencies to avoid going outside to a private contractor which will save both money and face in the long run. Many of the ethical questions can also be avoided. Along with the curation piece is location. Information professionals know where to look and, more importantly, how to look for information. This knowledge again saves time and money. The internet has useful intelligence data, but it is more useful if the searchers know where to search. Information professionals know where to dig. Reference librarians are especially skilled in this area so they bear an additional layer of responsibility.

There are some drawbacks to the entry of the library and information science field into the intelligence realm, however. Jin et al. points out that, while LIS and intelligence agencies already complement each other well, "LIS-oriented activities almost dominate collection and processing steps, while IA-oriented activities lead in the step to analysis (Jin & Bouthillier, 2012)." The take away from this is that information professionals do

not receive the skills to turn information into intelligence during their training. They are not a cure- all for the current system. If the intelligence community began taking students with information science backgrounds in after graduation there would be an increase in the number of courses dedicated to that subject matter. This is already happening in certain library programs across the nation. The University of Texas is a prime example of this. The Open Source Center of the CIA has gone so far as to set up an office in the main building of the School of Information.

The mentality of information professionals is slightly incompatible with the intelligence field. In this way they are tasked with bringing a different ethical viewpoint to the collection table. This mentality was made very apparent following September 11th. Government agencies requested the library records of patrons to analyze for potential threats to national security based on materials that had been checked out. In most cases librarians refused. Whether there is something about the library field that draws in individuals with this world view, or whether it is something learned throughout their course of study, there is a certain ethic that is instilled in library and information science professionals that runs counter to the traditional will of the intelligence community. This ethic is neither bad nor good; however, a counter stance to the norm is necessary for growth.

Conclusion

Open source intelligence has grown in use and usefulness as a result of advances in the information and communication technology sector. The advent of the internet has served to strengthen the bond between the two and will only function to increase the value of information gained from this type of collection for the foreseeable future. Open source intelligence is not without its problems. Social media is a question mark. It has the potential to be an excellent source, but legal issues, internal policy changes, and public opinion keep it constantly up in the air. As technology continues to advance so do the dangers that it can pose. China and the Middle East in particular raise a few problems. China for its content controls and the Middle East for its connectivity issues in Iraq, censorship in Iran, and decreasing penetration in Israel. Transparency and information overload make it a double edged sword. So do those who pose a threat to security be it national or global. The internet has become a weapon of information. Unlike the nuclear arms race, however, there is an upside. It is useful both for intelligence and everyday life. The ability for new technology to allow for many to many communication is something that OSINT needs to capitalize on, but in a delicate manner so as to avoid a situation of überveillance. Location data in the form of GPS is becoming so omnipresent that such an elaborate system is not necessary anyway. Policy needs to be put into place to clearly decide who, if anyone is responsible for data. The issues of transparency and information overload presented could be on their way to being fixed with the employment of librarians in curating and locating OSINT. With a few changes they could even become

proficient at analysis as well. The final take away from this is that open source intelligence and information and communication technology are always going to influence one and other for better or worse. The value of OSINT should not be underestimated because as time progresses the importance and influence of the internet is only going to increase. There are problems with OSINT as it relates to ICTs, but they are not insurmountable. If all goes well OSINT can become the dominant INT. Maybe then it will be HUMINT that will be the INT that becomes a side note in the intelligence process.

Bibliography

Chomik, A. (2011). Making Friends in Dark Shadows : An Examination of the Use of Social Computing Strategy Within the United States Intelligence Community Since 9 / 11. *Global Media Journal -- Canadian Edition*, *4*(2), 95–113.

Cosenza, S. V. (2012). World Map of Social Networks. *Vincos Blog*. Retrieved from http://vincos.it/world-map-of-social-networks/

Davies, P. H. J. (2002). Intelligence, Information Technology, and Information Warfare. *Annual Review of Information Science and Technology*.

Elgin, B., & Einhorn, B. (2006). The Great Firewall of China. *BusinessWeek*, 1–5.

Facebook. (2012). Data Use Policy. *Facebook*. Retrieved from https://www.facebook.com/legal/proposedDUP

Hartnett, S. J. (2011). Cyber-Warfare , Google , and U . S-Chinese Communication in an Age of Globalization. *National Communication Association*, pp. 1–2.

Hosenball, M. (2008). Cyber Spying For Dummies. *Newsweek*, p. 6.

Jin, T., & Bouthillier, F. (2012). The Integration of Intelligence Analysis into LIS Education. *Journal of Education for Library and Information Science*, *53*(2), 130–148.

Kessler, S. (2011). Mexican Blog Wars: Fourth Blogger Murdered for Reporting on Cartel. *Mashable*. Retrieved from http://mashable.com/2011/11/10/mexico-blogger/

Mercado, S. C. (2003). Sailing the Sea of OSINT in the Information Age. *Studies in Intelligence*, *48*(3), 45–56.

Michael, K., & Michael, M. G. (2007). From Dataveillance to Überveillance and the Realpolitik of the Transparent Society. *The Second Workshop on the Social Implications of National Security*.

Miniwatts Marketing Group. (2010a). Iraq - Internet Usage and Market Report. *The Internet Coaching Library*. Retrieved from http://www.internetworldstats.com/me/iq.htm

Miniwatts Marketing Group. (2010b). Iran - Internet Usage and Market Report. *The Internet Coaching Library*. Retrieved from http://www.internetworldstats.com/me/ir.htm

Miniwatts Marketing Group. (2012a). Internet World Stats. *The Internet Coaching Library*. Retrieved from http://www.internetworldstats.com/top20.htm

Miniwatts Marketing Group. (2012b). Isreal - Internet Usage and Market Report. *The Internet Coaching Library*. Retrieved from http://www.internetworldstats.com/me/il.htm

O'Connell, K. M. (2004). The Role of Science and Technology in Transforming American Intelligence, 14–26.

Pew Internet & American Life Project. (2012).*Pew Research Center*. Retrieved from http://pewinternet.org/Static-Pages/Trend-Data-(Adults)/Whos-Online.aspx

Protalinski, E., & Blue, V. (2012). The Social Web: Who owns your data? *ZDNet*. Retrieved from http://www.zdnet.com/debate/the-social-web-who-owns-your-data/10087130/

Resnyansky, L. (2009). The Internet and the Changing Nature of Intelligence. *IEEE Technology and Society Magazine*, 41–47.

Robbin, A., Courtright, C., & Davis, L. (2005). ICTs and Political Life. *Annual Review of Information Science and Technology*.

Schroeder, D. A. (2011). *Efficacy and Adoption of Central Web 2 . 0 and Social Software Tools in the U . S . Intelligence Community* (pp. 1–46).

United States Joint Chiefs of Staff (2000). Department of Defense Dictionary of Military and Associated Terms Includes US Acronyms and Abbreviations. *Department of Defense*.

Strickland, L., Baldwin, D. A., & Justsen, M. (2006). Domestic Security Surveillance and Civil Liberties. *Annual Review of Information Science and Technology*, 433–513.

Tait, R. (2006). Censorship fears rise as Iran blocks access to top websites. *The Guardian*. Retrieved from http://www.guardian.co.uk/technology/2006/dec/04/news.iran

Thompson, C. (2006). Open-Source Spying. *The New York Times*, pp. 1–14.

Wagner, D., Lopez, M., Doria, A., Pavlyshak, I., Kostakos, V., Oakley, I., & Spiliotopoulos, T. (2010). Hide and Seek : Location Sharing Practices With Social Media. *MobileHCI'10* (pp. 55–58).