

THE ROLE OF OSINT COLLECTION IN INTELLIGENCE AND THE FUTURE OF INTELLIGENT SEARCH

Elena-Adelina ANDREI

“Mihai Viteazu” National Intelligence Academy, Bucharest, Romania

Abstract: *In the Intelligence process, OSINT collection has a crucial role. Collection is the key to getting a good and useful product. Regardless of the field of study or preference for defining data - but especially in the Intelligence field - accurate data collection is essential to maintaining the integrity of the analytical product. The world is changing with the advancement of technology. Online sources prove more useful in gathering required intelligence to help analysts to understand some things. Nowadays, the Internet plays an important role in the OSINT collection, maybe the most important, because it provides lots of data that can be spread very fast. Therefore, the Internet is important for the growth of social media, media sharing sites, and their ease of access through various devices. Search engines are also experiencing a permanent development. For that reason, they provide a real support to analysts and their work. For instance, an eloquent example is the new Google service called Knowledge Graph. The Knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic search information gathered from a wide variety of sources. That way, it provides detailed information about a topic in addition to a list of links to other sites. That means users can use the information without having to navigate to other websites and assemble the information themselves.*

Keywords: *OSINT collection, Internet search tools, Knowledge Graph*

1. INTRODUCTION

Intelligence has increasingly become an activity based mainly on analysis of information on the use of open source and academic information. Consistent with these changes, modern intelligence means the ability to solve problems and bring new situations, process information, collect information, including to fully benefit from scientific, technological, military, economic, and political information. Today, OSINT holds an important place in the intelligence services' activity. With the passage of time and the evolution of technology, OSINT's contribution to the process of intelligence has changed. Open source development is largely due to the Internet and features that it provides. The explosion of intelligence information we are currently facing has been greatly magnified by what is called New Media, considered the main communication platform of the 21st century.

Society has evolved a lot and, at the same time, ways to collect information have greatly diversified. As stated by Wiewiorka & Wolton (1987), as not long ago it was very difficult to get informed because of the lack of information and poor access to it, today it is difficult to be properly

informed because of the high information flow. The most important moment in the OSINT development was the terrorist attacks of September 11, 2001 in the United States of America. After those events, it was realized the importance of open source information, corroborated with intelligence obtained from other sources, but especially the importance of information sharing.

Open sources were long disregarded before their importance has been recognized. Today, great emphasis is laid on open source information relevant to national security. Currently, a growing number of government departments and non-governmental organizations are actively exploiting open sources of information. Organizations and institutions operating in the fields of security, defense, and public order, which are in charge with ensuring state security, are increasingly cooperating with the public through the virtual world.

2. COLLECTION IN THE INTELLIGENCE PROCESS

In the simplest way to say it, the intelligence cycle is the process of developing unrefined data into polished intelligence for the use of customers. The intelligence cycle consists of six steps. The

second step, collection, contains both acquiring information and provisioning that information to processing and production components. Actually, collection represents the gathering of raw information taking into account requirements. Also, the collection process involves many activities, such as interviews, technical and physical surveillances, human source operation, searches, and liaison relationships (FBI.gov).

Collection strategies shifted from sophisticated surveillance satellites capable of counting tanks and missiles to the gathering of whatever information was available on rogue states and terrorist groups. Much of the information found was actually openly available but not hitherto collected and analyzed (CRS Report for Congress, 2007). OSINT collection is normally accomplished through monitoring, data-mining, and research. Open-source production supports all-source intelligence and the continuing activities of the intelligence process (generate intelligence knowledge, analyze, assess, and disseminate) OSINT is derived from the systematic collection, processing, and analysis of publicly available, relevant information in response to intelligence requirements. Two important related terms are open source and publicly available information: open source and publicly available information. OSINT collection is normally accomplished through monitoring, data-mining, and research (Headquarters, Department of Army, 2012). The collecting intelligence about individuals, groups, or states of interest has come under increasing scrutiny since September 11, 2001. Collectors have a specific view of the collection function, structure and process. Analysts need to take a different perspective, one that is not closely tied to the existing functional or structural divisions.

Collection operations depend on secure, rapid, and reliable communications to allow for data exchange and to provide opportunities for cross-cueing of assets and tip-off exchanges between assets. Once collected, information is correlated and forwarded for processing and production (Federation of American Scientists, 1996).

3. OSINT COLLECTION PROCESS

Intelligence collection requirements are developed to meet the needs of potential consumers. Based upon identified intelligence, requirements collection activities are given specific tasks to collect data. These tasks may use a number of different intelligence disciplines for collection

activities and the use of different types of collection systems contributes to redundancy. It also allows the collection of different types of information that can be used to confirm or disprove potential assessments. Several intelligence disciplines are used to acquire information, including human intelligence/HUMINT, open source intelligence/OSINT, imagery intelligence/IMINT, signals intelligence/SIGINT, measurement and signatures intelligence/MASINT (FAS, 1996).

Open source collection becomes more of a threat as more information becomes electronically accessible especially with the evolution of the Internet. This represents one of the risks of OSINT collection. As many others things, OSINT collection includes many risks and protective countermeasures should be developed to prevent inadvertent compromise of program activities through publication of data in publicly available media (Federation of American Scientists, 1996). Collection services include online collection - searchers that specialize in Internet, deep web and premium commercial online source exploitation (NATO Open Source Intelligence Handbook, 2001). Also, WebCase simplifies the challenge of documenting Internet based OSINT by allowing you to collect only the high quality information you need. Whether a news article, blog rumor, recruitment video, or “chatter” among tweets or forum postings – among many other forms of information – WebCase enables analysts to focus your intelligence gathering efforts for each case you examine (Vere Software, 2011).

Specialists recommend to analysts to establish the validity of their collected intelligence at the moment they collect it - even if the source web location is changed or deleted at any time afterward. This can be important when it comes to convincing their supervisors of the veracity of their actionable intelligence. Collecting information from open sources is generally less expensive and less risky than collection from other intelligence sources. The use of OSINT may result not only in monetary savings but also in less risk than utilizing sensitive technical and human sources (CRS Report for Congress, 2007).

4. THE GOOGLE KNOWLEDGE GRAPH

The Google Knowledge Graph (GKG) is a system launched by Google in May 2012, starting in the United States. Since July 2013, this system has been available for Romania, too. It can understand facts about persons, places and things

and how they are connected. The Knowledge Graph is used to enhance Google's search engine's search results with semantic search information added from a variety of sources (Singhal, 2012). It provides structured and detailed information about the topic in addition to a list of links to other sites. The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves (Waters, 2012).

According to Google team, the information in the Knowledge Graph is derived from many sources, including the CIA World Factbook, Freebase, and Wikipedia. The feature is similar in intent to answer engines such as Ask Jeeves and Wolfram Alpha (Staff, 2012). As of 2012, its semantic network contained over 570 million objects and more than 18 billion facts about and relationships between different objects that are used to understand the meaning of the keywords entered for the search (Isidoro, 2013). The Knowledge Graph is used both behind-the-scenes to help Google improve its search relevancy and also to present Knowledge Graph boxes, at times, within its search results that provide direct answers. Below are recent stories about the system (Search Engine Land, 2012). When it comes to search, we are accustomed to queries that are initiated client-side and not server-side. This new Knowledge graph update will essentially play an important role in how Google's engine understands your search queries and how you should SEO your website. The basic concept works as follows: when you type a phrase into the search box, like say for instance "Zama Zama", Google needs to know if you are searching for news about illegal mining, how to do the zama dance, or if you want to book tickets for Zama Zama the movie, which are completely different things. Once you type your particular phrase into the search box, Google will display all the related forms of your search on the right hand side, with small descriptions about each different interpretation of Zama Zama. You can then choose which interpretation you meant to search for and once you click on the particular field in the knowledge graph, your search results will refresh to show you the related information and search results (Zululex, 2012).

The second part of this update will then adapt the sidebar on the right to give you quick, factual information regarding your newly defined and more relative quire. For example, if you chose Zama Zama the movie, Google will display regular sites that mention the film in the search results, and

in the right hand sidebar quick booking options and cinema times will be displayed. Would you have searched for the director of zama zama Vickus Strijdom, a wikipedia like wiki will be shown in the sidebar with a short biography about Vickus. Google has a database with the facts people search for about Vickus the most, and will thus show only the information for these most popular facts. These summaries will contain links just like a wikipedia article, which will allow you to view summaries of related people, co-workers, family and other projects (Zululex, 2012). In the acceptance of Google's Vice President Ben Gomes, "To put this into perspective, one of the biggest launches in the past was Universal Search". He added that "the Knowledge Graph features affect a larger fraction of queries than Universal Search. Users will see these features more often than they see Google Maps in Google Search."

Google's latest upgrade will also deliver a brief profile of the queried subject, from an author's birthplace to a band's discography. Google use Knowledge Graph to showcase biographical details about a specific architect on its own website, serving up his birthdate, spouse, children and images of his most famous buildings. Google will customize the selection of facts to reflect users' most common queries relating to that person or place. The change puts Google front and center and promises to keep some users on Google, rather than diverting them to the websites where they could have found similar facts. It also promises to annoy some sites, which could see traffic diverted from their own pages, as well as attract the attention of antitrust regulators, who are already investigating charges that Google cooked its algorithms to feature its own products ahead of competitors (Bosker, 2012). The summaries will also add some surprises into users' search results, surfacing information users didn't know they wanted but may be interested in. The change underscores the steps tech companies are taking to engineer serendipity and introduce us to ideas we might not have discovered otherwise. Also, Ben Gomes mentioned that

We can connect things to what you're searching for in ways you might not have expected. In terms of search, this will greatly increase the amount of serendipity in search as you browse from a topic to things related to it.

Another new feature uses the Knowledge Graph to help users move from an ambiguous query. Google will present a list of more specific

queries to the right of its results. Gomes explains that Google hopes its Knowledge Graph will help the search engine answer more advanced questions that take into account specific locations and times of year, or require bringing together information from discreet databases (Bosker, 2012).

4. CONCLUSIONS

The collection and the use of open source intelligence (OSINT) can be of immense value to an analyst. Open source information may be indistinguishable from classified information, but based on its source, may not be evaluated as trustworthy. The classic OSINT collection process is about to change with GKG, this major upgrade to Google's search results, which aims to eradicate ambiguity and make clicking on the sites shown in search results increasingly unnecessary. Google notes that by connecting billions of disparate points of data, its search results can present users with a more coherent narrative, thus providing a better understanding of the relationships between facts. This is considered to be the future of intelligent search.

5. ACKNOWLEDGMENT

This article has enjoyed the support of the Project Pluri and Interdisciplinarity in doctoral and postdoctoral program cofinanced by Ministry of National Education – or Posdru, Contract no. POSDRU/159/1.5/S/141086. This paper is made and published under the aegis of the Research Institute for Quality of Life, Romanian Academy as a part of programme co-funded by the European Union within the Operational Sectorial Programme for Human Resources Development through the project for Pluri and interdisciplinary in doctoral and post-doctoral programmes Project Code: POSDRU/159/1.5/S/141086, Sectoral Operational Programme Human Resources Development 2007-2013. The contents of this material do not necessarily represent the official position of the European Union or the Romanian Government.

BIBLIOGRAPHY

1. Bosker, B. (2012). *Google Knowledge Graph Could Make Clicking Unnecessary*. [online]. Available: http://www.huffingtonpost.com/2012/05/16/google-knowledge-graph_n_1521292.html. [15.04.2015].
2. Brown, A. (2012). *Get smarter answers from the Knowledge Graph from Português to Google*. [online]. Available: http://insidesearch.blogspot.ro/2012/12/get-smarter-answers-from-knowledge_4.html. [25.04.2015].
3. CRS. (2007). Open Source Intelligence (OSINT): Issues for Congress. *Congressional Research Service*. [online]. <http://www.fas.org/sgp/crs/intel/RL34270.pdf>. [15.04.2015].
4. FBI. (2015). *Intelligence cycle*. [online]. Available: <http://www.fbi.gov/about-us/intelligence/intelligence-cycle>. [15.04.2015].
5. FAS. (1996). Intelligence collection activities and disciplines. *The Interagency OPSEC Support Staff*. [online]. <http://fas.org/irp/nsa/iOSS/threat96/part02.htm> [15.04.2015].
6. Headquarters, Department of Army. (2012). *Open Sources Intelligence*. Army Techniques Publication. [online]. Available: <http://fas.org/irp/doddir/army/atp2-22-9.pdf>. [15.04.2015].
7. Isidoro, A. (2013). *Google's Knowledge Graph: one step closer to the semantic web? Econsultancy*. June.
8. NATO. (2001). *NATO Open Source Intelligence Handbook*. [online]. Available: http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NA TO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf [15.04.2015].
9. Search Engine Land. (2012). *Google Launches Knowledge Graph To Provide Answers, Not Just Links*. [online]. Available: <http://searchengineland.com/google-launches-knowledge-graph-121585>. [15.04.2015].
10. Singhal, A. (2012). *Introducing the Knowledge Graph: Things, Not Strings*. [online]. Available: <http://googleblog.blogspot.ro/2012/05/introducing-knowledge-graph-things-not.html> [25.04.2015].
11. Vere Software. (2011). *How Authentication, Collection and Reporting Strengthen OSINT's Value*. [online]. Available: http://veresoftware.com/Using_WebCase_to_Support_OSINT_Collection-Final.pdf. [15.04.2015].
12. Waters, R. (2012). *Google To Unveil Search Results Overhaul*. *Financial Times*. [online]. <http://link.ft.com/r/QQSDPP/NJ7WZJ/1O1JA/7AEL7B/C48EYB/B7/h?a1=2012&a2=5&a3=17>. [25.04.2015].
13. Zululex. (2012). *Google launches its Knowledge graph*. [online]. Available: <http://seo.zululex.com/google-launches-its-knowledge-graph>. [15.04.2015].