OSINT: A MAJOR SOURCE OF UP-TO-DATE INFORMATION

Ephraim Lapid[1]

Introduction

In recent years, OSINT (Open source Intelligence) has undergone a quantitative and qualitative revolution. In the present article, we shall discuss recent developments in OSINT, (in particular Big Data Analysis) its use in Israel, as well as trends and challenges for the future. OSINT is the term for the processing of open sources information, and is therefore a meta-level. OSINT refers to intelligence information that has undergone integration, analysis, and evaluation. OSINT is the output of open source material processing.

Big Data Analysis

Big data analysis is a method of processing and constructing scalable data architecture for specific purposes, such as time series forecasting, hybrid intelligent techniques, and above all for implementation in cases of crucial decision-making. As its name suggests, the object of such analysis is to obtain throughput for big datasets, which exceed the capability of traditional data processing. Big data analysis also involves putting together data computing systems that facilitate high performance, efficient and scalable data mining for current use, and storage of data for future use

[1] Brigadier General (Res.) Ephraim Lapid, Ph.D., is a former Senior Intelligence Officer in Israel.

The differences between big data and traditional data are summarized by 3V's – volume, velocity, and variety. Any category of big data, e.g. geo-physical data, comprises by definition large-volume datasets. In certain cases, these datasets exceed the advanced computational capabilities achieved over the past years. A typical example is the datasets used by Google Earth, which have yielded an archive of over 20 petabytes of imagery, comprising satellite images, aerial photos and ground-level street view descriptions or imagery. Likewise, Google Street View, which projects a car driving down a street, is equipped with numerous cameras and sensors, capturing millions of image pixels per second, while simultaneously scanning thousands of point locations to come up with clear images.

Velocity does not merely refer to the rate at which the data is generated or produced, but also to the currency of its content and the corresponding need for timely analysis. In terms of velocity, big data as a whole has reached a volume that exceeds computational capabilities over a given time.

Variety refers to the diversity of sources and types of data being processed, and to the varied degree of extraction and distribution of data from such sources. This notable diversity regards the types, extractability and distribution of data from such sources. Over the years, the price of data storage, particularly on internet applications, has gone down. Filtering generally involves reduction of data attributes, data cleaning, data integration, and data transformation. The end goal of filtering is to polish the data derived from open source such as the internet, thus making them exploitable for purposes such as intelligence or policy-making.

Falsification

How reliable, valid and relevant are sets of data derived from open sources, particularly from the internet? The answer would depend, in many cases, on the degree of falsification and the attitude of the producers, the individuals participating in the creation of data. It would also depend on whether the information they produce is true or false. Their internet behavior would also depend on whether any entity i.e. their state authorities, their friends and family or their employers (has put them under control). The issue of falsification is therefore crucial for intelligence professionals but no less so, for policy-makers, who must ascertain that the data on which they base their policy are reliable, relevant and validated. Although OSINT has been hailed as a relatively inexpensive intelligence discipline, falsification of open source information on the Internet has become a true challenge in terms of OSINT data usability. In most cases, the law-enforcement agencies, through diverse intelligence operations, are the ones to inform people that they are being placed under control online. This, in turn, triggers falsification, thus generating highly unreliable, unfounded and worthless bulks of information on the internet. In many instances, individuals on their own initiative create false identities or avatars to avoid detection or other reasons. One technical solution that has been proposed to counter the increased probability of falsification is using computational models based on “trust scores” of social networks, to determine the trustworthiness of relevant information .

Social Media on the Internet: Major Sources of OSINT, or a Major Problem?

Although the internet is not the only source of OSINT, it has been acknowledged as its major source. In the extreme, people who use social media tend to create their own version of their life by feeding in enhanced and tweaked information about themselves. At the same time, this implies that the volume of activity on the Internet, and hence the corresponding data and information used and stored, increases by gigantic amounts, making navigation and filtering of the relevant pieces of information considerably difficult .

Approximately 2.5 billion gigabytes (2.5 petabytes) are generated on the internet every day, constituting the raw bulk from which data can be mined. Is this amount of information a blessing or a curse?

Data mining, also known as knowledge discovery, is the process of extracting meaning from a large amount of data. Given that massive amounts of data are available on the internet, data mining is a necessary tool to obtain relevant intelligence information. This is difficult albeit doable, and still less costly than resorting to other forms of intelligence information gathering .

OSINT in Israel

Israel, as many other countries, gathers intelligence information primarily on enemy military capabilities (such as ORBAT, activity, combat means and deployment), as well as on political issues. The information is used to formulate an intelligence assessment of the intentions and activities of the adversary. In the past decade, the information collection effort has undergone a significant change, and it currently focuses primarily on non-state enemies, terror organizations and individual terrorists, who constitute the main threat against numerous states around the world. “The masses now have the floor”. In intelligence jargon this phenomenon is known as as "Mass Perception" has resulted in, increasing the weight of social media surveillance .

Consequently, OSINT has been thriving. The web and the "global village" era have added strength to new information resources other than the traditional mass media. These transformations have confronted the intelligence services with the need to prepare for the production, understanding and analysis of a new world of content and technology, and handle hundreds of thousands of new types of information units every day.

New technological capabilities had to be developed to filter out and classify the unprecedented volumes of information originating in social networks, blogs, web forums and Wiki sites. The diversity of formats and languages, work in a virus-contaminated environment, the need to overcome defense mechanisms and impersonate human surfers – all these have yielded impressive intelligence capabilities. Nowadays, the intelligence value of the daily report issued by the OSINT unit is much higher than before. With the internet revolution, open sources have mushroomed and their output has become more diversified. These enormous quantities have called for new solutions – technological and content-related – that would generate significant metadata and enable processing huge quantities of information in seconds.

A key side effect of the information-explosion age is a boost in information visualization. Intelligence consumers are currently offered OSINT products in the form of video clips that illustrate the situation more vividly than the old textual reports. They generally prefer dedicating a few minutes to watching an integrative visual intelligence product to reading thousands of documents.

In addition to their intelligence value, open sources also play an important role in national propaganda efforts ("Public Diplomacy"). In recent years, efforts have been made worldwide to justify and establish the legitimacy of every operational measure taken to thwart the resolute and manifest objectives of terror organizations.

Future Challenges of OSINT

1. On-demand information: Providing the users with content adjusted to their areas of interest. Websites and content providers have been studying the users' needs, and now use the push technology to provide them with relevant content and sources. This mechanism is already highly developed on Twitter, where the surfers are even able to air their feedbacks on the proposed content on the spot. By indicating the extent of the content's relevance to them, they help channel to them relevant information only.

2. Constant and significant development of visual content search tools (pictures and clips): Developing highly advanced search engines that enable the surfers to locate visual data and access visual content channels according to their own definitions. If in the past a picture could be located only be the textual descriptive details that accompanied it, Google and other search engines are currently able to retrieve similar or even identical pictures or video clips based on their visual content.

3. The world of information on smart phones: This is a true revolution that has made redundant the traditional mass media – the printed press, and radio and television news. Smart phones are in fact palm-size open source gathering agencies. They facilitate an ongoing consumption of content adjusted to the users' definitions.

4. Twitter: A highly valuable source of news and other information on any spot on the globe and any sphere of interest. This is also the preferred medium of journalists, who use to twit their reports, including rich and varied visual information, long before they appear online or in print.

5. Breaking the language barrier: Automatic translation has been advancing rapidly. Translation engines currently allow basic reading of texts in any language, including exotic ones. Automatic translation tools are easy to use (outrageously so, in

the eyes of linguists), but the beauty of this point is that a good professional translator is able to improve the output of an automatic translation engine, and produce a highly accurate product rapidly and effectively. These engines keep improving and are likely to become almost 100% accurate eventually .

6. Higher information reliability: Computational models based on “trust scores” of social networks could prove to be the technical solution against the growing probability of falsification. Sequential or temporal pattern mining can also be used in order to establish the reliability of Internet sites. Sequential pattern mining would involve identifying consistent clustering and order-preserving regularities of occurrences of data in internet sites .

7. Further development of automatic tools to adjust the information to the preferences of individual users.

8. Intensified efforts to design information security means and solutions given the increasing threats against individual privacy and against business, national and security information in the cyber space.

In summary, OSINT has become a developing information-collection agency throughout the modern world of intelligence. Intelligence bodies and individuals keep expanding their use of information, profiting from the various technological security and civil applications now available in cyberspace. Professional collaborations are recommended, as they would significantly enhance the filtering and processing of open-source information in this age of metadata.