OPEN SOURCE INTELLIGENCE (OSINT). THE WAY AHEAD

Gabriel-Traian UNGUREANU

"Mihai Viteazul" National Intelligence Academy, Bucharest, Romania

Abstract: *The technological advances have led to an exponential growth in the amount and complexity of open-source information. The weight of this type of information in the collection base has significantly increased and now we have intelligence services that collect more than 80% of their intelligence from open sources. Therefore, in this article, referring to state-of-the-art technologies in an exhaustive approach, we analyze how the new technologies sustain the intelligence services in their actions for collecting open-source intelligence and increase the efficiency of related processes. Moreover, we analyze how these technologies can supplement and enhance one another throughout the phases of intelligence cycle to render the intelligence services' activity more efficient. Finally, analyzing the historical progresses, we will foretell what technologies are expected to support OSINT in the next ten years and how they will be integrated into the phases of intelligence cycle, forecasting how OSINT will advance through technological change.*

Key words: *open-source intelligence, OSINT, technological change, intelligence cycle, new technologies impact.*

1. INTRODUCTION

The process of collecting open-source information is considered and approached differently, depending on the field of activity. From one area to another (academic, business, journalistic or national security intelligence), the strategic vision, the related operations or the tactical steps are distinct when it comes to open sources. The main difference of OSINT – open-source intelligence concerning national security – is the large amount of data and information collected. Another distinguishing feature of OSINT is the huge number of sources, extremely varied sources which also require various collection methods. The mechanisms used by the government intelligence services to process open source data into validated information (from OSD - Open Source Data into OSIF - Open

Source Information into OSINT - Open Source Intelligence into OSINT-V - Open Source Validated Intelligence) are also highly complex (NATO: 2011).

Despite the above-mentioned difficulties, the intelligence services are currently able to provide actionable intelligence collected only from open sources. Also, OSINT can provide support or contextual intelligence. Thus, together with Human Intelligence (HUMINT), Signals Intelligence (SIGINT), Imagery Intelligence (IMINT) and Measurement and Signature Intelligence (MASINT), the intelligence services can ensure, by integration and all-source analysis, the information required by the beneficiaries (NATO: 2011). Along with the other types of sources, OSINT significantly contributes to the collection base.

In order to address the challenges posed by OSINT and not only, due to the volume of data and information, the number and diversity of sources and the complexity of specific processes of intelligence cycle, the intelligence services have always innovated and used state-of-the-art technologies (CRS: 2020).

As the purpose of this article is to estimate what technologies will enhance OSINT in the next ten years and how these technologies will be integrated into the specific processes of information collection, processing, analysis, dissemination, and storage, in section II "OSINT" we will analyze the evolution of OSINT from the early modern period to the present days, establishing the current and future trends. In section III, considering the trends of OSINT and the possible technological advances, we will analyze the technologies that are currently used and will foretell what technologies are to be used in the near to medium future.

2. OPEN SOURCE INTELLIGENCE (OSINT)

2.1. OSINT Evolution

Even if the earliest historic references on open-source collection date from ancient times, its specific processes were defined once the print media appeared.

The first relevant landmark retained when discussing OSINT evolution is the start of systemic and systematic collection from open sources. In their fight for control of the Mediterranean Sea, the Venetian Republic and Dubrovnik Republic (or Ragusa) significantly contributed to the development of intelligence. Thus, in the 15th century, both Venetia and Ragusa held structured (systemic) intelligence networks which were constantly and systematically transmitting information. In the context of this fierce fight for information collection, at the middle of the 16th century, both Venetia and Ragusa

noted "the value of the gazettes that started to circulate" (Huges-Wilson: 2018, 29). This is why the first systematic collection from open sources appeared. "Once the print media appeared, the espionage became more open and systematic". Moreover, "the Venetian agents started to collect market information when the big European fairs were organized" (Huges-Wilson: 2018, 30). As such, around the year 1540, in order to leverage the circumstances promoting the collection of open and semi-open-source information, Venetia had in place an emerging (systemic) network of intelligence about money, markets, trade and the related flows, which was being used in the State's interest (Huges-Wilson: 2018).

The second landmark in OSINT development is the moment when a structured intelligence network influenced and controlled the print media. In 1865 a secret power center was set up in the service of Otto von Bismarck: a network of more than 45,000 spies coordinated by Wilhelm Stieber. The network provided domestic and foreign information and often carried out missions against diplomatic targets. For example, for the 1870 attack against France, Stieber organized the largest intelligence system involving more than 35,000 people and providing more than 1,650 reports for each possible and probable issue (Huges-Wilson: 2018). Moreover, he first started a genuine psychological war with the aim of improving his own army morale and weakening the enemy's morale by publishing the enemy's errors, bad news and losses and emphasizing their own success. Besides the political, diplomatic, or military aspects, it is also relevant to mention that Stieber "was the first head of a national intelligence service that used agents to monitor and control the press" (Huges-Wilson: 2018, 46). From that moment, the intelligence services had to validate the information collected from the press based on criteria such as publication, author, context, etc.

The third landmark is the moment when the intelligence services were turned into government institutions and expanded. Due to the social tensions arisen as a result of the actions carried out by the secret services, the intelligence services became a public obsession and the governments, fueled by these fears, started to set up the first national organizations. Thus, the Secret Service Bureau was formed in 1909, changing subsequently its name in MI5. In 1912 the "Secret Intelligence Service" was formed and was in charge of foreign operations. It subsequently became MI6 (Williams, Blum: 2018) (Herman: 1999). The development of domestic and foreign intelligence services has implicitly led to the development of OSINT.

The approach was now bivalent, the open-source information being collected from domestic sources and from abroad target territories (Herman: 1999).

The fourth landmark would be the rise of radio and television. During the two world wars, intelligence acquired a substantially increased importance. In the US, the Office of Strategic Services (OSS), the CIA predecessor, was founded. Stalin also developed the secret services. In this context of continuously expanding intelligence services, the need for OSINT development became apparent (Herman: 1999). As the Cold War was an intelligence war, all types of sources had to be accessed and all known collection methods were used, OSINT playing an increasingly important role, also considering that this period coincided with the development of public and private networks of radio and television. (Williams, Blum: 2018). While the intelligence services were previously focused on indexing, storing and easily accessing written information, they started now to be concerned with converting voice to text in order to process the information disseminated through radio. This is also the moment when emerged the need to capture, index, save and access video information. It is the period in which it became apparent that, in addition to SIGINT, IMINT or MASINT, the efficiency of OSINT largely depended on the efficiency of computing and other technologies used (Williams, Blum: 2018).

The fifth landmark is the mass use of the Internet. The fact that the start of www coincided with the end of Cold War, the removal of certain international tensions and the eradication of some censorship systems led to the exposure and transit of a much higher volume of data, also due to the speeding up of globalization. The year 2000 and the dotcom bubble burst caused an exponential growth in the volume of data available on the Internet (CRS: 2020). Moreover, social media opened up new perspectives, providing new categories of data and meta-data which allowed the micro-profiling and profiling of an entity or a group at any level (Miller: 2015) (Huges-Wilson: 2018). This trend was further enhanced by the emergence of mobile data access devices which led to a significant increase in the use of social media (Williams, Blum: 2018). The new technologies provided for the intelligence services' needs of managing huge volumes of data, facilitating the specific activities of the intelligence cycle, especially as concerns the collection, processing, management and dissemination of information.. For example, in 2008, DNI Open-Source Center (OSC) was using technology to collect economic information from

over 2,000 periodicals, 300 radio stations, 235 television stations, 95 foreign organizations (e.g. The Economist, Thomson Reuters, Lexis-Nexis, Stratfor, etc.), and other sources from more than 160 countries, with information in more than 80 languages, from websites, press, shows, television, radio, maps, databases, grey literature, photographs and commercial satellite imagery (Hamilton: 2011). Moreover, from the early 2000s, the new technologies have allowed for the implementation of efficient communication flows and procedures, in real time, between various groups of entities within the intelligence structures, also complying with the restraints imposed by information classification criteria (ieee.es: 2013). The moving from analog to digital made it possible to fine tune the collection methods. Radio streams were automatically converted into text and the television streams into text and frame sequences. This enabled their automatic indexing and processing, for example the who-where-when correlations. The search and indexing by image became possible for all categories of data. The correlations OSINT – IMINT – SIGINT – MASINT were also enabled etc. At this point, the challenges posed by OSINT consisted in the ability to convert into actionable intelligence the large volumes of data, which in most cases were unorganized, came from multiple sources, were available in different forms and collected through several categories of channels and to subsequently convert them into validated intelligence (Hamilton: 2011). The specific methods and instruments could assist in the creation, validation and enhancement of a collection base adjusted to the intelligence needs. It is an essential step considering the high number of sources, of different types (databases – government and private – structured, unstructured databases, studies, prospects, publications, etc.), coming from various key areas (economics, politics, diplomacy, state administrative apparatus, etc.) with anisotropic topics (processes, products, regulations, competitors, mergers, partnerships, sectoral information, social trends, etc.) (Brown: 2019). The calibration of collection base is a process where the sources are continuously reviewed, keeping however in mind that the sources should cover all key areas required for the intelligence base.

Summing up the five landmarks, we briefly present below the evolution of OSINT in the past 200 years:

• the advent of print media leads to the establishment of intelligence systems allowing the systemic and systematic collection from open sources;

• the structured intelligence networks start to influence and control the print media and the intelligence services need to categorize the information collected from the press based on criteria related to publication, author and context;

• the intelligence services become government institutions and the segregation between domestic and foreign occurs, the open-source information being collected now from homeland sources and from abroad target territories by different structures;

• public radio and television networks are developed and, if until that time, the intelligence services were focused on indexing, storing and easily accessing written information, they start now to be concerned with converting voice to text in order to process the information disseminated through radio. The need to capture, index, store and access video information becomes also apparent;

• in addition to SIGINT, IMINT or MASINT, the efficiency of OSINT starts to highly depend on the efficiency of computing equipment and technologies used;

• the massive use of the Internet causes an exponential growth in the volume of data available on the Internet;

• the moving from analogic to digital allows for the segmentation of data and data series, facilitating the collection, processing, and analysis;

• the social media and the emergence of mobile data access (www) devices provide new categories of data and meta-data which allows for the micro-profiling and profiling of an entity or a group at any level.

• using the new technologies, the OSINT structures transform into actionable intelligence big volumes of data, which are most often unorganized, derived from multiple sources, available in different forms and collected through several categories of channels;

• as the volumes of available data are huge, the purpose is no longer to collect all information that could be of interest, and the collection base needs to be continuously adapted to the intelligence needs.

Considering the above-mentioned landmarks, we identify the following trends:

1. the volume of open-source information is continuously growing at an exponential rate;

2. the number of sources is continuously increasing, and their types are more and more diverse;

3. the technological progress and the new technologies allow for the automated collection and processing of large volumes of data and their integration from various types of sources;

4. the processes for setting source system limits and calibrating the collection base in relation to intelligence needs are dynamic.

2.2. States, Strategies and OSINT Centers

In the light of the above-mentioned trends, we can infer that OSINT is placed high on the agenda of most intelligence services. This is further reflected in the strategy of intelligence community and is part of the national security strategies.

As such, the National Security Strategy of the United States of America - 2017, in a first mention concerning OSINT, emphasizes that the US rivals (with direct reference to China and Russia) use marketing information and techniques to attack individuals and institutions. It is emphasized that the risks to US national security grew as the attackers integrate information collected from multiple sources and use methods of analysis based on artificial intelligence tools. The countermeasures identified by the American government concern the collection and enhancement of information from all available sources and creation of cutting-edge technological platforms. Moreover, within one of the priority strategic actions in terms of intelligence, the US "will, in concert with allies and partners, use the information-rich open-source environment to deny the ability of state and non-state actors to attack our citizens, conduct offensive intelligence activities, and degrade America's democratic institutions" (White House: 2017, 32), and this strategic line is based on the new technologies. The idea is detailed in the chapter "Information Statecraft" which clearly mentions the tools used: large databases integrating information derived from personal and commercial sources with open-source information and data analytic and processing capabilities based on Artificial Intelligence (White House: 2017).

The French Defense and National Security Strategic Review - 2017 is focused in Part C on the strengthening of its Defense Industrial and Technological Base (DITB). This technological infrastructure sustains the national economy and helps France to expand its influence over the world. France's strategic autonomy and technological superiority are possible only if it sustains and maintains the excellence of entities within the DITB. Furthermore, one of the approaches supported by the Ministry of Defense requires that the administration, in partnership with the research community, should make a systematic use of open sources (President of the French Republic: 2017).

In the Russian National Security Strategy - 2009, the innovation and investments in the new technologies are referred to as national strategic priorities. Moreover, clear strategic lines, such as the development of applied sciences, of new technologies or the interdisciplinary cooperation, are detailed (Russian Federation's President: 2015). Even if there are no direct references to open-source

collection, the contextual references are frequent (Russian Federation's President: 2015) (CRS: 2020). Similar references are also found in "Doctrine of Information Security of the Russian Federation" – 2016, which makes mention of technical intelligence (The Ministry of Foreign Affairs of the Russian Federation: 2016).

In the "National Security Strategy - 2013", Japan underlines its need to strengthen the capabilities of intelligence community. Chapter 7 "Enhancing Intelligence Capabilities" clearly states that the development of OSINT is crucial, because in addition to HUMINT, SIGINT, IMINT and other types of sources, it will ensure a consistent system of sources (Ministry of Foreign Affairs of Japan: 2013). As a sign of strategical consistency, in "2020 Defense of Japan", OSINT is referred to as being one of the six pillars that contributes to the development of the capabilities of intelligence services (Ministry of Defense of Japan: 2020).

Similar approaches are found in the security strategies of other states, such as Canada, England, Germany, the Netherlands or China, the latter also displaying a special focus on OSINT (Bereziuk: 2016).

Thus, we may conclude that open-source collection is among the main concerns of the intelligence services all over the world.

As the importance of open-source collection has increased over time in terms of volume, being reflected accordingly in the composition of the collection base, the intelligence services were required to use to a greater extent the technology in the process, in order to collect and automatically process the avalanche of large data volumes (Tabatabaei, Wells: 2016). As such, dedicated structures have been formed and strengthened, such as "Open-Source Center" within the CIA (USA), "Centre des sources ouvertes" within the DNI (France) or "The Internet Center" within the NPA (Japan). Russia acts through the 12th Chief Directorate within the GRU, "information warfare", FAPSI "The Federal Agency for Government Communications and Information" and IRA "Internet Research Agency". China established "the Third Department" or "Technical Department of the Central Military Commission".

3. TECHNOLOGIES

While before the end of Cold War the technologies used mainly originated from the research centers within the defense and security structures, in the past 20 years the academic and business environments have also significantly contributed to the technological change, as more and more joint projects and programs were implemented (Williams, Blum:

2018). One such example is Hacking for Defense, which currently involves DOD, IC, 31 universities and 66 companies. These programs gave rise to a new trend, namely: if until the 1990s most of the cutting-edge technologies were shifting from the defense to the business environment, after 2000 this trend reversed. Moreover, private companies acting in the competitive intelligence field started to be noticed (Katz: 2020). As such, now we can make a relatively thorough assessment of the technologies used in OSINT with reference to unclassified information and information preponderantly coming from open sources.

Thus, considering the phases of intelligence cycle, we list below, in table I, the new technologies which could ensure a more efficient exploitation of intelligence in OSINT centers (Williams, Blum: 2018).

Table 1. OSINT Cycle and Technologies

OSINT Cycle	Operational Steps	Intell Category	Technologies							
			CSS	BD	AI			RPA	BC	Δ
					SA	ML	NLP			
Collection	Scanning Sources	OSD	CSS		SA		NLP			Δ
	Monitoring Sources		CSS		SA	ML	NLP			Δ
	Collection		CSS							Δ
	OSD Storage			BD	SA			RPA		Δ
Processing	Normalize	OSIF OSINT			SA	ML	NLP			Δ *1
	Aggregate				SA	ML		RPA		Δ
	Processed Data Indexation				SA		NLP			Δ
	Correlations				SA	ML	NLP			Δ
	Clearance Permissions				SA	ML		RPA		Δ
	OSIF Storage			BD	SA			RPA		Δ
Analysis	Source Validation	OSINT OSINT-V			SA					Δ *2
	Intell Validation				SA	ML	NLP			Δ
	Integration			BD	SA	ML	NLP	RPA		Δ
	Analyze				SA	ML	NLP			Δ
Dissemination	Production	OSINT-V			SA	ML	NLP	RPA		Δ
	Intell Indexation				SA	ML	NLP	RPA		Δ
	Distribution				SA				BC	*3
	OSINT Storage			BD	SA			RPA		Δ
	Feed-back						NLP		BC	Δ

OSD:	Open Source Data
OSIF:	Open Source Information
OSINT:	Open Source Intelligence
OSINT-V:	Validated OSINT

Δ - other applications
*1 - convertors
*2 - social network analysis; geointell
*3 - electronic post applications

CSS:	Crawlers - Spiders - Scrapers
BD:	Big Data
AI:	Artificial Intelligence
ML:	Machine Learning
NLP:	Natural Language Processing
SA:	Statistic Algorithms and Analytics
RPA:	Robot Process Automation
BC:	Blockchain Technologies

Source: author, 2021

Further analyzing the above contextual framework, we can note that in the near future, the most significant impact on the economic intelligence systems will be that of the new technologies (Johnson: 2010) (ODNI, DNS: 2018). As "Table I" indicates, there are four large types of technologies that have and will have an increasingly important contribution: Big Data, Artificial Intelligence, Robot Process Automation and Blockchain.

3.1. Big Data

Big Data is the technological field that deals with systematic ways for storage, extraction and analysis of data and information. Big Data appeared as a solution to the problem raised by data sets that were too large or complex to be handled by traditional applications. Big Data is associated with three key concepts: volume, variety, and velocity (Miller: 2015).

Big Data facilitates the capturing, storage, update, querying, search, sharing, transfer, analysis and visualization of data, while ensuring the confidentiality of information and sources. Big Data makes it possible to identify correlations between specific data sets, to determine specific patterns, and promotes the use of predictive analysis and other analysis methods that allow for the extraction of actionable intelligence (Katz: 2020).

The high capacities of storage, indexing and analysis of information allow to easily access large amounts of data and information (Johnson: 2010) (ODNI, DNS: 2018). The databases can be connected to automatic systems for collection of open-source information, coupled with systems for automatic indexing of collected data and information, which enables the structuring of data and information, also including the implementation of policies regarding the permissions (the rights of access) (Miller: 2015). Moreover, 5G technologies will significantly increase the technical and tactical capabilities in the field of data collection, monitoring, telecommunications (extensive real-time data transfers from the tactical area – automation based on transmission via cellular networks, Internet of Things, or edge computing). At the same time, vulnerabilities specific to these capabilities will arise and require significant counterintelligence efforts, especially in the field of cyber security (The Economist – Intelligence Unit: 2020).

3.2. Artificial Intelligence

Artificial Intelligence is intelligence demonstrated by

technological devices that collect signals and perceive the environment where they operate, starting and managing actions that maximize the chances to successfully achieve their objectives, by replicating the cognitive functions associated to human mind, such as learning, analysis and problem solving (Miller: 2015).

Artificial Intelligence includes three broad types of algorithms:

- Statistical Algorithms and Analytics – generation of statistical structures and calculation of probability, predictive analytics,

- Machine Learning – neural networks and reasoning, deep learning, and

- Natural Language Processing – learning and representation, accessing of knowledge (Williams, Blum: 2018) (Nacci: 2017).

In most software solutions, these technologies supplement and enhance one another, processing and analyzing large databases (Big Data Analytics), their result being used and interpreted in classical software and programming structures (Williams, Blum: 2018) (Nacci: 2017).

The Artificial Intelligence technologies can allow the automation of collection platforms, also optimizing the selection of sources based on collection requirements. The algorithms can generate models that are able to anticipate certain collection tasks according to currently processed information, can trigger the selection of the best sources or determine optimal collection frequencies. Also, deep-learning algorithms facilitate automatic decision-making allowing for a dynamic adaptation of the collection tasks. Thus, the collection becomes adaptable also entailing a downsized human contribution (Katz: 2020) (Nacci: 2017).

Moreover, the artificial intelligence applications will allow the use of efficient automated solutions during the information processing phases and later in the analysis phase. Scenario analysis, predictive analysis, establishment of both recurrent patterns and future patterns are possible using artificial intelligence technologies (The Economist – Intelligence Unit: 2019). Furthermore, the development of hardware capacities (e.g. quantum computers) will enhance the capabilities of Artificial Intelligence-based software, having a significant future impact. According to a study issued by the Emerging Technology and National Security Team, 32% of critical technologies used by the intelligence systems will be based in the future on Artificial Intelligence and 16% on quantum computers (Miller: 2015) (ODNI, DNS: 2018).

3.3. Robot Process Automation

Robot Process Automation is the technology that allows to automate IT and digital processes by configuring software robots that emulate and integrate the actions of a human operator, interacting within the digital systems to carry out various processes. RPA robots can utilize the user interface to collect data and to manage applications, just like a human user. RPA robots interpret information, trigger responses, and communicate with other systems, especially in order to perform a wide variety of repetitive tasks (CRS: 2020).

Robot Process Automation technologies are useful for the processes of intelligence collection, processing, dissemination, and management because they allow the automation of recurrent tasks by processing and indexing large volumes of data, also ensuring the correlations between databases, recipients, channels of communication and other entities (Miller: 2015).

3.4. Block Chain Technologies

Block Chain Technologies are based on a chain of cryptographically linked data blocks, each block containing a cryptographic key ("hash") of the previous block, a timestamp and transaction data. Therefore, Block Chain Technology does not allow any data alteration. Once recorded, the data from a block cannot be retroactively altered without the alteration of all subsequent blocks. The technology uses a distributed ledger and not a unique ledger.

Block Chain Technologies will allow for an efficient implementation of the policies regarding the permissions and rights of access to classified information, including the remote access to systems and data on high security conditions (The Economist – Intelligence Unit: 2019).

3.5. Other Technologies

Web Scraping is used to extract data from websites. Using various protocols, the software directly accesses web pages through automated processes, a bot, or a Web Crawler, to collect specific data (from a database) which are stored and subsequently used (Pinto, et al: 2018). Web Scraping implies downloading the content of a page and then extracting data for subsequent processing. Thus, the content of a page can be stored, structured, and subsequently analyzed. The Web crawler, the Spider or the Spider-bot is a robot that systematically browses the web. Crawlers copy the pages for immediate or later processing. Crawlers can also map the web identifying the relevant information for specific search sequences.

All these technologies are essential for collection because they allow for the collection and storage of web data (Katz: 2020).

The analog-to-digital converter is a system that converts an analog signal (sound, light etc.) into a digital signal. A converter can also provide the digital value of a measurement, for example the value of voltage in a circuit. These technologies are critical for OSINT because they allow the conversion of analog information, such as audio recordings, video recordings or images, into digital information. The digital information has the advantage that it can be stored, indexed (attributes can be assigned to digital information), aggregated and subsequently used in intelligence actions. Some specific technologies of the intelligence community that allow certain digital-digital conversion required for an appropriate indexing of information can be included as well in the class of convertors (Williams, Blum: 2018). For example, the automatic conversion of an audio recording into text or the conversion of a digital video recording into audio and then into text and correlation of the text with the frames. Similarly, printed information can be converted into digital information. Also, using Artificial Intelligence technologies, subjects can be automatically identified in photos, as well as the actions taken, territories and other information which can be indexed, subsequently allowing a proper exploitation (Williams, Blum: 2018).

Social networks analysis software facilitates social network analysis in terms of quantity or quality, by describing network characteristics and by numerical or visual representation. The networks can consist of anything, from families, teams, disease vectors and membership on social media websites and web networks (Pinto, et al: 2018). Networks are formed of direct connections between nodes or indirect linkages based on shared attributes, shared attendance at events or common memberships. High performance software packages use relational databases to import and store some network features and use advanced Artificial Intelligence algorithms for statistical analyses, sentiment analyses, traffic, language and prediction (Katz: 2020).

The electronic mail applications are critical, because they are vulnerable mainly regarding data transfers security. Security requirements, such as the identification of the person and related device, data encrypting and security, network and data packages security, are critical issues that these applications must ensure. Also,

the applications can be configured so as to be in accordance with the distribution matrices, to add beneficiaries or, for certain topics, to suggest the user that he or she should also inform other entities that could face the risk (Brown: 2019).

4. OSINT - FUTURE TRENDS AND TECHNOLOGIES

We will refer below to OSINT by analyzing the phases of intelligence cycle: collection, processing, analysis and dissemination. Please note that our approach will be somehow theoretical, organized by individual phases. In fact, the intelligence cycle phases often overlap or in some cases, they are not all part of the process.

As regards all phases listed below, it is relevant to mention that for the entire intelligence cycle the matters pertaining to the safety of databases, devices and networks, data submissions and packages are central items on the agenda of intelligence services (Katz: 2020) (CRS: 2020) (Tabatabaei, Wells: 2016). In most cases, dedicated modules or applications are being developed, but efficient solutions have also been developed in the commercial sector, such as those provided by large companies like Microsoft, IBM, Cisco, Fortinet etc. or other niche providers, such as SolarWinds, Intruder, Bitdefender, Malwarebytes, etc.

4.1. Collection

As regards the collection, we first refer to the setting of the source system for all systematic (recurrent) intelligence topics. As such, based on the national interests, most intelligence services also constantly monitor other target territories. In many cases, the monitoring implies the setting of a system of sources, an initial correlation of relevant information available in those sources, followed by the systematic collection of new information. Such collection can be planned at various intervals or can take place at the same time as the publication. Once an optimal frequency is set, the sources are systemically and systematically scanned and monitored: new intelligence is gathered, the differences arising in databases are identified, volume or qualitative variations in the news flow are assessed or the deviations of certain indicators are recorded and flagged (Katz: 2020). As soon as the collection base is set and validated, the sources remain relatively stable, meaning that, once set up, the technologies used for collection operate without requiring other significant adjustments, unless the source structure is being adjusted (ODNI, DNS: 2018).

However, the appearance of asymmetrical topics can require a recalibration of the source system. Even though by the surveillance and monitoring activities, the intelligence services aim to be warned about, and timely anticipate all possible threats, atypical disruptive situations that had not been covered by those mechanisms can occur. Such atypical situations require that new sources be integrated in the collection base. The integration of new sources requires in most cases that reliable technological solutions be found as soon as possible. As such, it is recommended for the above-mentioned collection base to be able to ensure the informative framework capable to provide early warnings and respond to all types and categories of possible and probable threats (CRS: 2020).

Therefore, the recurrent collection of intelligence is a preponderantly technological, systemic, and systematic process, the purpose of which is to ensure an exhaustive information base. The collection of information by request is also a technological but non-systemic and non-systematic approach, aimed to ensure an information base for a specific topic. The ideal scenario for any OSINT center is to operate relying only on the systemic and systematic intelligence collection, but due to the volatility and asymmetries of the current environment, the professionals are often required to dynamically adjust the system of sources and need to find new technological solutions to handle specific requests for information (Katz: 2020).

While at present the sources that are preponderantly exploited are those allowing for a systematic collection, the technological advances will soon make it possible to map most of the open sources and to identify more easily technological solutions that enable the extraction of information from these sources. It is hence confirmed the trend according to which the processes for setting source system limits and calibrating the collection base based on information needs are dynamic (Brown: 2019). This will be possible because, with the exception of print media, there will be new technologies to enable an efficient identification and monitoring of web, media, social media and dark web sources (Ortega: 2019) (Pastor-Galindo et al. 2016). On the one hand, there are new solutions allowing the systemic collection from certain categories of sources and, on the other hand, the existing platforms are developed so as to include and process new categories of sources (Brown: 2019).

To identify the open sources from web, social media and dark

web, the OSINT centers currently use web search engine-like solutions that are somehow similar to Qwiki, Wolfram Alpha, Wowd or Recorded Future.

The future trends indicate the use of elastic search solutions, enhanced by robotic process automation technologies, solutions which would enable an efficient scanning of social media (Kanakaris et al., 2018). Another aiding tool is the specific artificial intelligence technology, which by special statistical algorithms, natural language processing and machine learning, allow a proper identification of sources. Moreover, the software solutions will run on continuously upgraded hardware infrastructures. The increase in the operating speed and computing power will promote the development of more and more complex and refined algorithms (CRS: 2020).

The best-known core robot process automation solutions are: UiPath, blueprism, Automation Anywhere, etc.

It is difficult to give examples of artificial intelligence solutions as their field of application is extremely wide. Large corporations, such as Microsoft, IBM, Google, Adobe, OpenAI etc., along with niche companies, such as Starmind, Vesttorly, Braina or Footprints are among the best-known providers.

To extract data, the ONSINT centres currently use "spider", "scrap", "crawl", "geolocation" or other similar solutions, enhanced by artificial intelligence and robot process automation capabilities (Williams, Blum: 2018) (Kanakariset al., 2018) (Pastor-Galindo et al., 2016). The solutions used by the OSINT centres are similar to solutions such as: Scrapy, HTTrack, OutWit Hub, Visual Scraper, Google Dorks, Spyse, Creepy, etc.

The future trends indicate the use of these types of solutions (Crawlers – Spiders – Scrapers) for data extraction, but within them, the weight of artificial intelligence and robot process automation will be significantly higher.

4.2. Intelligence Processing and Indexation

Both systematically and non-systematically collected (for a specific topic) information should be normalized, aggregated, indexed (assigning of attributes) and stored, so as to be easily identified whenever necessary (Katz: 2020).

Normalization means converting and bringing all collected information to the same form (e.g. all information should be in digital text format, in the same language and all images in a specific digital format) (Miller: 2015). The efficiency of

conversion software is critical here. Identical and similar data that is collected from different sources is then aggregated (Williams, Blum: 2018). Subsequently, in the indexing phase, along with the allocation of metadata and attributes (subjects, places, events – subject matters, actions, thematic verticals), depending on the technologies used, the related potential risks and recommended countermeasures can also be correlated, or other similar cases can be listed, etc. (Katz: 2020) (Pastor-Galindo et al., 2016). The more attributes and metadata are allocated, the higher the likelihood of leveraging the information at a subsequent phase or operation. For these categories of applications, it is relatively difficult to identify examples of commercial software (Williams, Blum: 2018). However, certain features are found in applications such as CRMs or MATLAB, SPSS, Wolfram Mathematica, etc. With regard to databases, at the forefront are the solutions provided by large corporations such as IBM, Oracle, Microsoft, etc. or for the analytical part, IBM SPSS, Elasticsearch, SAP Hana, MicroStrategy, SAS Visual Analytics, Forestpin Analytics, etc.

Moreover, the access rights can be technologically set up to a great extent, in keeping with the policies on information classification. For example, the rights can be allocated by correlating the levels of access to information with those of a subject and with access levels that have been allocated to the sources and by subsequently fine-tuning permissions based on the metadata related to the information or the cluster where the information was distributed (Katz: 2020). Here we can also find solutions developed by Oracle, Microsoft, Cisco, etc.

Besides the aspects related to the indexing and setup of access rights, the processing and indexing of available web and dark web databases are other challenges posed at this point. Difficulties arise as each source can use arbitrary structures and indexing for its databases and the normalization, aggregation and indexation thereof require great initial effort (Ortega: 2019). For example, a simple parameter, such as the distance between two points, can be expressed in kilometers or miles. Or a source can rely on the distance between the two points, while other source can relate to an average time needed to cross that distance (Katz: 2020). There are cases where maintenance can also be difficult because the sources can voluntarily change the structure of the bases. Moreover, it would be ideal for all databases collected from various sources and relating to a certain topic

to be aggregated into a single base. It is only after normalization and aggregation that the warning and early warning indicators (levels) can be determined. The aggregation and normalization precede the automated and semi-automated generation of recurrent informative products submitted to certain beneficiaries. An example of informative products that are automatically issued based on series of data are those reflecting the fluctuations of currencies in certain territories (Katz: 2020).

The current trends indicate that the processing will be preponderantly carried out using technology. The advances in conversion applications, the fine-tuned normalization and aggregation solutions assisted by robot process automation will allow the specific artificial intelligence (machine learning and natural language processing) algorithms and technologies to successfully perform these categories of tasks.

4.3. Analysis

A first challenge in OSINT analysis is the validation of information. Besides the actual analysis of information, attention should be paid to sources and authors in order to avoid all possible attempts of manipulation or misinformation (Nacci: 2020). There are publications or authors who get voluntarily involved in manipulation or misinformation and authors who have unwillingly been trapped into the manipulative arrangements, unknowingly taking over certain distorted ideas in their articles (Katz: 2020). Moreover, incomplete information and fragmented information are hindrances that can be however compensated at a subsequent phase, by analyzing all sources (by also integrating other types of sources – HUMINT, IMINT, SIGINT, MASINT, etc. in the analytical approach) (Katz: 2020) (Brown: 2019). For information analysis, the Artificial Intelligence solutions based on natural language processing and machine learning become more and more efficient, making it possible to detect any attempted manipulation, propaganda, or fragmentation of information (Brown: 2019) (Nacci: 2020). In this respect, commercial solutions, such as IBM i2, Palantir, Siren, Cogito, 6th Sense, etc., are in place.

The second challenge in OSINT analysis consists in determining intelligence dissemination patterns, given that information is often taken over without quoting the source. In the analytical phase, it is essential to determine the source of information and the time of publication, because possible interests related to the source or author can also be analyzed

(Katz: 2020) (Brown: 2019). This operation is even more difficult in case of social media and can be in all likelihood carried out only using high-performance technological tools (Williams, Blum: 2018) (Kanakaris; et al., 2018). The impact of artificial intelligence and robot process automation is increasingly felt in this case too. At present, the machine learning algorithms and robot process automation mechanisms make it possible in certain cases to automatically identify the source (Miller: 2015) (Brown: 2019). In parallel, social network analysis applications, such as NetMiner, NetworkX, Centrifuge, SocNetV, etc. have also been developed.

The impact of Artificial Intelligence will be noted in case of analysis too. The correlations who, what, why, where, how, with whom, etc. will be more and more accurate, providing the analyst with significant clues (Brown: 2019). The analogies related to certain patterns of action, profiling, speech analytics, evidence management, etc. will be processed automatically to a greater extent, providing the analyst with reliable working assumptions (Katz: 2020). A significant role in this respect will be played by quantum computers, which will accelerate the development of artificial intelligence and revolutionize certain programming patterns (Miller: 2015).

4.4. Intelligence Dissemination

In case of OSINT, the creation of informative products gives rise to several situations. As such, certain informative products can be automatically created (using technology), with a specific pre-set frequency. Examples of such informative products include the press review or a summary report or chart on stock exchange transactions concerning the companies in which the state holds equity interests. Likewise, in certain cases the reports are automatically generated, and the analyst conducts a final review of the document before being submitted to the beneficiary (Nacci: 2020). A third case is that in which the informative products are fully drafted by intelligence analysts, in keeping with the common practices, in most cases according to an approach implying the analysis of all sources (Katz: 2020).

With regard to intelligence indexing and classification, all documents are indexed and stored according to the principles set forth under the processing phase, because it is highly important to be able to further access them whenever necessary.

In terms of intelligence dissemination, the informative products are generally disseminated to the designated beneficiary or

beneficiaries. However, the current technologies facilitate the creation of distribution matrices, which makes it possible to also disseminate this information to other beneficiaries (Katz: 2020). The distribution matrices make it possible to draw up the best lists of beneficiaries, taking into account the permissions to access the information and the interests of beneficiaries. Depending on the metadata attached to the informative products, the informative products are automatically correlated with other possible beneficiaries who may leverage such information, which significantly enhances the efficiency of this informative approach (Williams, Blum: 2018).

The beneficiaries may subsequently request clarifications or additional information. They may also submit suggestions regarding the presentation form (text, table, charts, images, etc.), calibrating hence the communication to the best level (Williams, Blum: 2018).

There are numerous applications designed for electronic mail services, including here players such as Microsoft, IBM, Apple and other niche players, such as eM, Inky or Hiri. As to the next steps, the trends of an increasing impact of Artificial Intelligence and Robot Process Automation are once again confirmed here. Moreover, block chain technology plays a significant part when it comes to transaction security (Brown: 2019). Here too there are well-known providers, such as IBM, Microsoft, Oracle, etc. but also smaller companies, including Ripple, NEO, Stellar, etc.

5. THE NEXT STEPS OF OSINT

The already identified five trends are further confirmed by the outcomes of each OSINT step, as follows:

1. The volume of open-source intelligence is continuously growing, at an exponential rate, a trend confirmed by the development of web, dark web and social media.

2. The number of sources is continuously increasing, and their types are more and more diverse, a trend which is supported both by the diversification of technologies used in web and databases, as well as by the emergence of content platforms.

3. The technological advances and the new technologies allow for the automated collection and processing of large volumes of data and their integration from various types of sources, a trend sustained by the technological advantages related to Artificial Intelligence and Robot Process Automation.

4. The processes for setting source system limits and calibrating

the collection base in relation to the information needs are dynamic and can be technologically assisted by Artificial Intelligence algorithms.

5. OSINT collection and processing progresses towards a preponderantly technological approach, as confirmed by each phase of the intelligence cycle. As such, trend 5 can be reworded as follows: OSINT progresses towards a preponderantly technological approach.

5.1. The merge between OSINT and the new Technologies

Hence, OSINT progresses towards a technological approach. Each phase uses its own technologies, Crawlers – Spiders – Scrapers in case of Collection, convertors in case of Processing or Blockchain in case of dissemination. It is relevant to mention that Artificial Intelligence and Robot Process Automation contribute to all phases. Moreover, the advances made in the development of quantum computers, the increased computing power, the continuous development of various types of networks will significantly influence the collection, while also prompting and providing solutions, which will also bring about some direct benefits, such as a better processing speed, an increased volume of processed data and improved ways of processing and analyzing information (Katz: 2020). Likewise, quantum encryption and decryption could radically change this field. Quantum computers could accelerate the development and increase Artificial Intelligence capabilities, transforming the collection and processing capacities (Nacci: 2020). The development of 5G technology and IoT devices will trigger changes in where and how information is collected, the "omnipresent connectivity", creating new opportunities for the collection field (Katz: 2020) (Nacci: 2020).

In conclusion, the technological advances support the development of OSINT. Practically, the accelerated technological development inevitably leads to an accelerated development of OSINT.

5.2. The OSINT impact in the national intelligence systems

The sustained growth of OSINT will reflect accordingly in the weights of collection base and subsequently, in those of the intelligence base. The advantages provided by the automation of collection, processing and dissemination phases and the significant technical support provided to the analysts enhance the importance of open-source collection. Moreover, the fact that information is collected remotely, without exposing the human collector, is a significant

advantage (Ivanjko, Dokman: 2020).

Therefore, due to the increase in the volume of open-source intelligence, due to the fact that OSINT processes are automated, reducing hence significantly the required resources and, as the open-source acquisition does not expose the human collectors, OSINT will continue to play a central part within the ecosystems of intelligence services.

5.3. The Next Steps

We can conclude that in the near future the ONSINT cycle will be a preponderantly technological approach.

As regards the collection, within the next ten years the technological advances will make it possible to map most open sources and identify easily, often automatically, the technological solutions for extracting information from these sources. The processes for setting source system limits and calibrating the collection base in relation to the information needs will be automated because, except for print media, the technological solutions will allow for an efficient identification and monitoring of web, media, social media, and dark web sources. The future trends indicate that Crawlers – Spiders – Scrapers solutions will be used for data extraction, but within them, the weight of artificial intelligence and robot process automation will be significantly higher. Moreover, "elastic search" solutions will be used and will be enhanced by robot process automation technologies, leading to an efficient scanning of social media. Artificial intelligence will promote a proper identification of the sources.

Processing will be preponderantly carried out using technology. The conversion applications, the fine-tuned normalization and aggregation solutions assisted by robot process automation technologies will enable the specific artificial intelligence technologies and algorithms to successfully carry out these categories of tasks.

Intelligence analysis will rely on efficient natural language processing and machine learning algorithms, which will automatically detect any attempted manipulation, propaganda, and fragmentation of intelligence. The correlations who, what, why, where, why, with whom, etc. processed by Artificial Intelligence will be more and more accurate. The analysis of scenarios, analysis of concurrent hypothesis, speech analytics, evidence management, etc. will be processed to a great extent automatically, providing the analyst with reliable

working assumptions. The network analysis applications enhanced by machine learning algorithms and robot process automation mechanisms will lead to an automate identification of the source.

The dissemination will use advanced protocols of encryption and secured transfer of information, probably using quantum computing. Moreover, the distribution matrices will allow the dissemination of information to optimized lists of beneficiaries.

6. CONCLUSIONS

How the intelligence services will adjust and synchronize OSINT with the technological process will be an essential success factor. Four levels are to be noted here:

- Use of artificial intelligence technologies in the specific phases of intelligence cycle and in most specific software platforms;
- Use of robot process automation technologies in the specific phases of intelligence cycle and in most specific software platforms;
- Adoption of 5G technologies;
- Adoption of quantum technologies.

These levels could make the difference between an efficient and effective service and a lower performance service.

ACKNOWLEDGMENT

This article was supervised by Professor Irena Chiru, PhD.

REFERENCES

[1] Bereziuk, B., (2016) *The Modus Operandi of Chinese Intelligence*, Ottawa: Carleton University, pp. 3-6.

[2] Brown, Z. T., (2019) *Adaptive Intelligence for an Uncertain Age*, Washington: National Intelligence University, pp. 15-188.

[3] Hamilton, B., (2011) *No More Secrets - Open Source Information and the Reshaping of U.S. Intelligence*, Santa Barbara: Praeger Security International, pp. 3-84.

[4] Herman, M., (1999) *Intelligence Power in Peace and War*, Cambridge: Cambridge University Press, pp 259-268.

[5] Huges-Wilson, J., (2018) *Serviciile Secrete*, Bucharest: Meteor Publishing, pp. 29-46.

[6] Ivanjko, T., Dokman T., (2020), *Open Source Intelligence (OSINT): Issues and Trends*, Zagreb: University of Zagreb, pp 6-9. gate Publishing Company, Farnham, pp. 304-306.

[7] Johnson, L. K., (2010) *National Security Intelligence*, Oxford: Oxford Handbooks, pp. 12-14.

[8] Kanakaris, V., Bandekas, D.V., Tzovelekis, K., Geo-Location on Twitter and Instagram Based on OSINT Techniques: a Case Study, In: *International Journal of Advanced Research*, February 2018, pp. 780-789.

[9] Katz, B., (2020) *The Collection Edge*, Washington: CSIS Briefs, pp. 2-8.

[10] Miller D. T., (2015) *Defence 2045*, Washington: Rowan & Littlefield, pp. 21-52.

[11] Nacci, G., (2017) *Appunti sulla architettura sistemica delle F.onti in OSINT*, ResearchGate, Working Paper, pp. 7-10

[12] Nacci, G., (2020) *OSINT e COVID-19*, Rome: Società Italiana di Intelligence Press, pp. 04-07.

[13] Ortega, J. M., (2019) *OSINT + PYTHON: Extracting information from TORnetwork and Darkweb*, Alicante: University of Alicante, pp. 12-77.

[14] Pastor-Galindo, J., et al, (2016), *The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends*, Murcia: University of Murcia, pp. 6-8.

[15] Pinto, R. A., Hernández Medina, M. J., Pinzón, C. C., Díaz, D. O., García, J. C., Inteligencia defuentes abierta (OSINT) para operacionesde ciberseguridad. Aplicación de OSINT en un contexto colombianoy análisis de sentimientos, In: *Revista Vínculos: Ciencia, Tecnología ySociedad*, vol 15, n° 2, julio-diciembre 2018, pp. 197-201.

[16] Tabatabaei F., Wells, D., (2016) *Open Source Intelligence Investigation - OSINT in the Context of Cyber-Security*, Heidelberg: Springer International Publishing AG, pp. 215-221.

[17] Williams, H. J., Blum I., (2018) *Defining Second Generation OSINT for the Defense Enterprise*, Santa Monica: RAND Corporation, pp. 1-14.

[18] ***CRS Report, (2020) *Artificial Intelligence and National Security*, Washington: Congressional Research Service, pp. 2-29.

[19] ***ieee.es - Spanish Institute for Strategies Studies, (2013) *Economic intelligence in a global world*, Madrid: Spanish Ministry of Defence - Strategic Dossier 162, pp. 118-119.

[20] ***Ministry of Defense of Japan, (2020) *2020 Defence of Japan*, Tokyo: Ministry of Defense, p. 23.

[21] ***Ministry of Foreign Affairs of Japan, (2013) *National Security Strategy*, Tokyo: Ministry of Foreign Affairs of Japan, p. 18.

[22] ***NATO, (2011) *NATO Open Source Intelligence Handbook*, SACEUR, pp. 2-4.

[23] ***ODNI, DNS, (2018) *Emerging Technology and National Security*, Washington: US Government, pp. 14-27.

[24] ***President of the French Republic, (2017) *Defence and National Security Strategic Review*, Paris: President of the French Republic, pp. 63-72.

Russian National Security Strategy, Moscow: Russian Federation President, pp. 6-28.

[25] ***Russian Federation's President, (2015).

[26] *** The Economist – Intelligence Unit (2019), *A Whole New World: How technology is Driving the evolution of intelligent banking*, London, pp. 3-25.

[27] ***The Economist – Intelligence Unit, (2020) *A strategic C-suite playbook for navigating the 5G world*, London, pp. 10-12.

[28] ***The Ministry of Foreign Affairs of the Russian Federation, (2016) *Doctrine of Information Security of the Russian Federation*, Moscow: The Ministry of Foreign Affairs of the Russian Federation, pp. 5-12.

[29] ***White House, (2017) *National Security Strategy*, Washington: White House, 2017, pp. 32-35.