Open-Source Intelligence - Useful Tools in Data Analysis

Adelaida STĂNCIULESCU
Bucharest Court, Bucharest, Romania
adelaida.stanciulescu@gmail.com

Abstract

The paper aims to address how open sources, available in the public space, can provide relevant, high-quality information on which organizations (whether public or private) can strengthen their decision-making process. For example: the development of public policies, the development of security policies, law enforcement norms, the adaptation of tax systems to the digital age, the implementation of targeted marketing campaigns, the widespread access to continuing education, with the aim of creating an adapted workforce in the digital age, the business environment can support technology change through a more intense collaboration with authorities, local communities and society as a whole, etc.

Index terms: Open Source Data (OSD), Open Source Information (OSINF), Open Source Intelligence (OSINT)

1. Introduction

In this article I aimed to present the fundamentals of Open Source Intelligence (OSINT), how it is used, as well as the tools and techniques that can be used to collect and analyze information from open source (Open Source Data). In the era of globalization and digitization, information has become the resource without which progress, at this moment, seems impossible. In this context, the analysis of information from open source (Open Source Data), available on the web, has become a requirement, and even a necessity.

Starting from the premise that information is the first and most important element of the decision, we understand its importance and applicability in all fields, from political to military, from social to economic and even cultural [1]. Over the past two decades, the entire world has witnessed profound transformations as a result of globalization and technological change.

The galloping evolution of information and communication technology, recorded in the last two decades, has opened up new opportunities aimed at significantly improving the techniques and methods that can be used in data analysis.

Using Open Source Data (OSD) competitive intelligence analysis, organizational leaders can gain a valuable perspective and engage in debates to find beneficial solutions in a world of more and more virtual interactions. Before presenting the sources and use of Open Source Intelligence (OSINT) it is important to understand the terms used in this field.

2. Terms used in the field of Open Source Intelligence (OSINT)

2.1. Open Source Data (OSD)

According to US public law [1], Open Source Data (OSD) is publicly available information from open sources. According to the OSINT Guide developed by the Romanian Intelligence Service

[2], data from Open Source Data (OSD) are considered information communicated through radio/TV broadcasts, prints, unprocessed signals, photographs, tapes, satellite images and personal letters.

Although the definition published in the OSINT Guide seems more rigid, in my opinion it is much more clarifying, because it is accompanied by the source of this data, thus, as can be seen, the data from the open source Open Source Data (OSD), can be considered those data which are:

- Published or broadcast in the public space (for example, news media content);
- Available to the public on request (for example, census data);
- Available to the public by subscription or purchase (e.g. industry magazines);
- Could be seen or heard by any casual observer;
- Made available at a meeting open to the public;
- Earned by visiting a place or participating in any event that is open to the public.

From both perspectives on the definition of the notion of Open Source Data (OSD), the public nature of this data undoubtedly follows, this information is "publicly available". The term "open source" refers specifically to information that is available to the general public. If specialized skills, tools or techniques are required to access information, it cannot reasonably be considered open source. Paradoxically, open source information (OSD) is not limited to what we can find using the main search engines.

Web pages and other resources that can be found using Google are certainly massive sources of open source information (OSD), but they are far from the only sources. According to former Google CEO Eric Schmidt, a huge proportion of the information available on the Internet, more than 99%, cannot be found using the main search engines. This so-called "deep web" is a mass of websites, databases, files and more, which (for a variety of reasons, including the presence of login pages or barriers raised by payment mechanisms) they can be indexed by Google, Bing, Yahoo or any other search engine. Despite this, much of the content of the deep web can be considered open-source data (OSD) because it is readily available to the public.

Such legal, open sources (so-called "white sources") include, but are not limited to:

- National business registers;
- The official documentation that companies must present by law, including financial statements;
- Information from relevant state offices and units (Public Procurement Office, Chief Inspectorate for Environmental Protection, Consumer Protection Office, etc.);
- Bankruptcy notices in court and information from debt exchanges;
- Statements of spokespersons for companies and state persons;
- Press and mass media;
- Social networks;
- Social surveys;
- Public life.

2.2. Open-Source Information (OSINF)

Public legislation in the US does not make a strict delimitation of the notions: information produced from open source data (OSD), i.e. Open Source Information (OSINF) [2] and Open Source Intelligence (OSINT), letting them complement or substitute in places, while the OSINT Guide developed by the Romanian Intelligence Service [2], draws clear limits regarding this notion as follows, by Open Source Information (OSINF) - we mean: correlated and processed data to create information of general interest - articles from mass media, books, communiques.

2.3. Open Source Intelligence (OSINT)

As a complex, specialized and distinct process, Open Source Intelligence (OSINT) integrates human experience, with data obtained from open sources, in order to produce information and informative documents relevant to the decisions of leaders, regardless of the type of organization they belong to.

According to US public law [1], Open Source Intelligence (OSINT) is:

"(1) ...information produced from open source data collected - Open Source Intelligence (OSINT) , exploited and disseminated in a timely manner to an appropriate public, in order to respond to a specific information requirement."

"..."

"(3) Open-source intelligence production is a valuable intelligence discipline that must be integrated into the tasks, collection, processing, exploitation, and dissemination of information to ensure that United States decision makers are fully and completely informed."

According to the OSINT Guide developed by the Romanian Intelligence Service [2], by Open Source Intelligence (OSINT) we mean the results of a complex OSD and OSINF processing process, which involves identification, validation of sources, collection, corroboration and analysis, in order to develop products with relevance in terms of national security, which correspond to specific intelligence requirements.

Open Source Intelligence (OSINT) uses advanced technology to discover and analyze massive amounts of data, obtained by scanning public networks, from publicly available sources such as social media and the deep web - content that is not crawled by engines but search, which is, however, publicly accessible.

OSINT tools can be open source or proprietary: a distinction must be made between open source code and open source content. Even though the tool itself is not open source, as an OSINT tool it provides access to openly available content known as open source intelligence [4].

OSINT is in many ways the mirror image of operational security, which is the security process by which organizations protect public data about themselves that could, if properly analyzed, reveal damaging truths. IT security departments are increasingly tasked with conducting OSINT operations within their own organizations to strengthen operational security.

2.4. Validated Open Source Intelligence (OSINT-V)

Another notion encountered in the field of Open Source Intelligence (OSINT) is: Validated Open Source Intelligence (OSINT-V). Thus, according to the same OSINT Guide, developed by the Romanian Intelligence Service [2] - by Validated Open Source Intelligence (OSINT-V) we mean those data which have a high degree of certainty, either because they are made by a professional analyst or because they come from reliable open sources.

As a partial conclusion, we can state that the four previously presented notions actually constitute the four stages completed in the Open source intelligence (OSINT) process.

So:

1. Open Source Data (OSD), represents the initial stage of collecting raw data from several different sources - without analyzing and processing this data;
2. Open Source Information (OSINF), the second stage, which consists of an initial grouping of the data collected during OSD and an initial general analysis of the information held;
3. Open Source Intelligence (OSINT), i.e. the transfer of processed data to the requester;
4. Validated Open Source Intelligence (OSINT-V), based on checking information already available in other open sources and sometimes comparing it with data obtained through other methods.

In the following we will focus on how Open Source intelligence (OSINT) can be used as best practices for cyber security. There are two common usage scenarios:

- Ethical hacking and penetration testing
- Identification of external threats

Ethical Hacking and Penetration Testing

Security professionals use open source intelligence to identify potential weaknesses in managed networks so they can be fixed before they are exploited by threat actors.

Common weaknesses include, but are not limited to:

- Accidental leaks of sensitive information, for example, through social networks;
- Open ports;
- Unsecured devices connected to the Internet;
- Software without updates, such as websites running old versions of common CMS products.

Identification of external threats

The Internet is a great source of information that can provide information about possible attacks on an organization, identifying new vulnerabilities and how they are being actively exploited, to intercepting threat actors' conversations about an upcoming attack.

Thus, Open Source Intelligence (OSINT) allows security professionals to prioritize their time and resources to address the most significant threats.

In most cases, this type of activity requires an analyst to identify and correlate data from multiple sources to validate a threat before taking action. For example, while a single threatening post on a social network may not be a cause for concern, the same post would be viewed in a different light if it were linked to a group of known threats that he is active in a certain community.

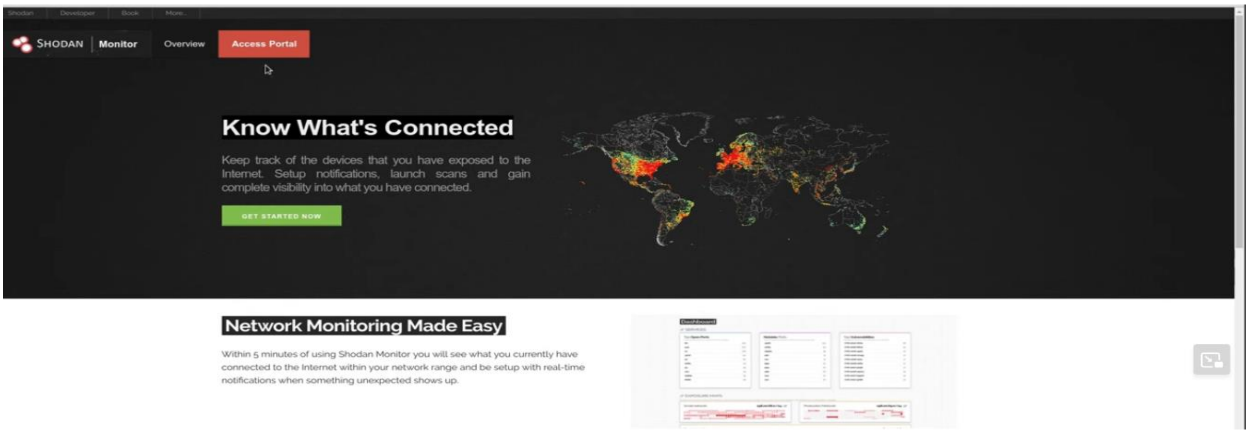
One of the most important things to understand about open-source intelligence is that it can often be used in combination with other subtypes of intelligence. Information from closed sources such as internal telemetry, dark closed communities and external information sharing communities are regularly used to filter and verify open-source information.

Case study on collecting data and turning it into useful information

As an analyst, having a large amount of information at your disposal is both a blessing and a curse. On the one hand, you have access to almost anything you could possibly need, but on the other hand, you have to be able to find what you need by actually searching through a torrent of data.

In this example we will use online tools other than traditional search engines. As you know, Google is the most used search engine, but Shodan is a search engine that provides results for security professionals and more, being a goldmine for hackers to see exposed assets.

Shodan is a security monitoring solution that makes it possible to search deep web and IoT networks. It makes it possible to discover any type of device connected to a network, including servers, smart electronic devices and web cameras. It mainly includes information related to the assets that are connected to the network. Devices can range from laptops, traffic signals, computers and various other IoT devices. This open-source tool mainly helps the security analyst to identify the target and test it for various vulnerabilities, passwords, services, ports and so on.



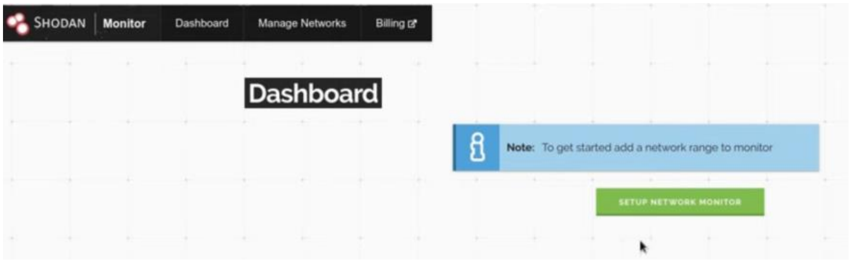
Know What's Connected

Keep track of the devices that you have exposed to the Internet. Setup notifications, launch scans and gain complete visibility into what you have connected.

Network Monitoring Made Easy

Within 5 minutes of using Shodan Monitor you will see what you currently have connected to the Internet within your network range and be setup with real-time notifications when something unexpected shows up.

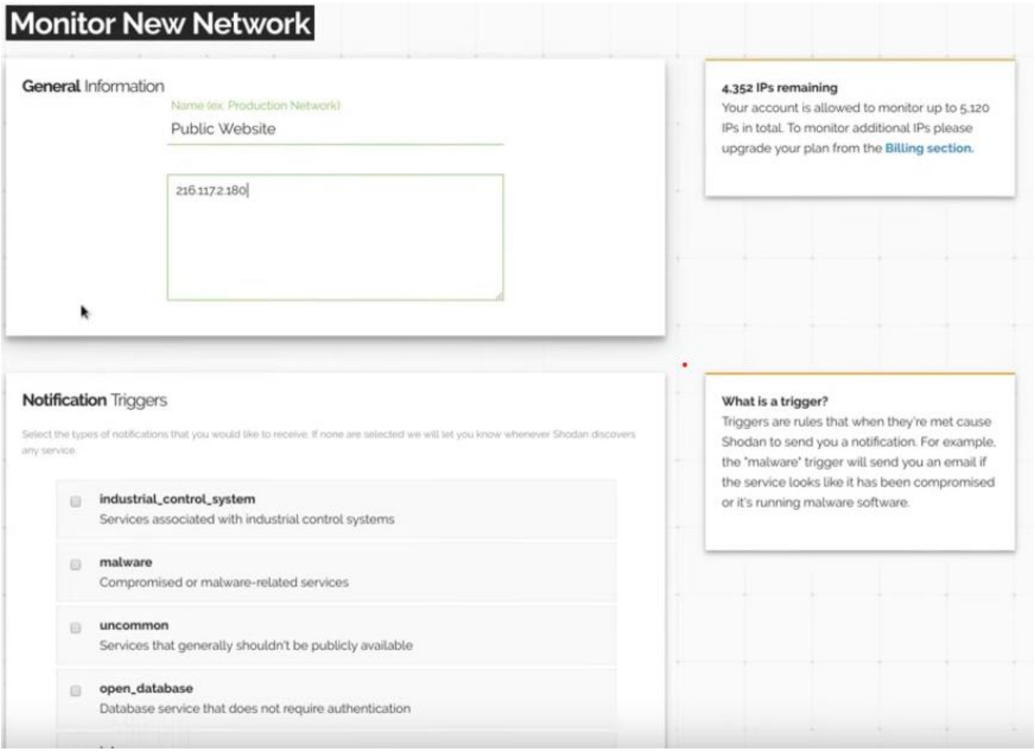
Shodan manages to identify and test "default passwords", devices with VNC viewer, use of open RDP port for testing available assets, etc. This tool is available at https://www.shodan.io and requires login to access the information. After authentication, the network to be monitored must be configured.



Dashboard

Note: To get started add a network range to monitor

In this example, we configure for monitoring a site located at the IP address 216.117.2.180.


Monitor New Network

General Information

Name (ex: Production Network): Public Website

216.117.2.180

4,352 IPs remaining
Your account is allowed to monitor up to 5,120 IPs in total. To monitor additional IPs please upgrade your plan from the **Billing section**.

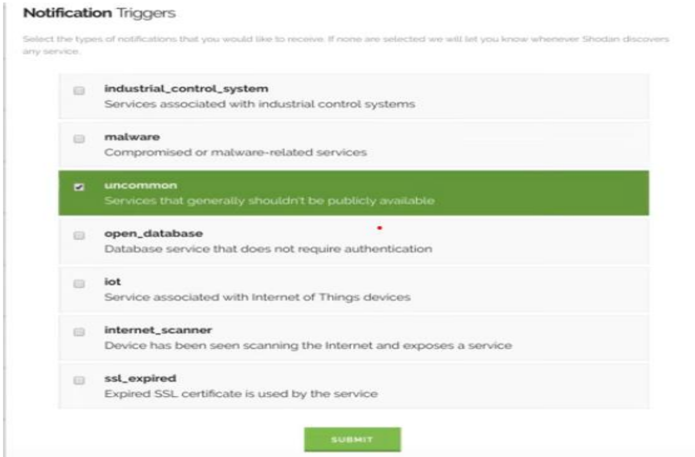
Notification Triggers

Select the types of notifications that you would like to receive. If none are selected we will let you know whenever Shodan discovers any service.

- □ **industrial_control_system** Services associated with industrial control systems
- □ **malware** Compromised or malware-related services
- □ **uncommon** Services that generally shouldn't be publicly available
- □ **open_database** Database service that does not require authentication

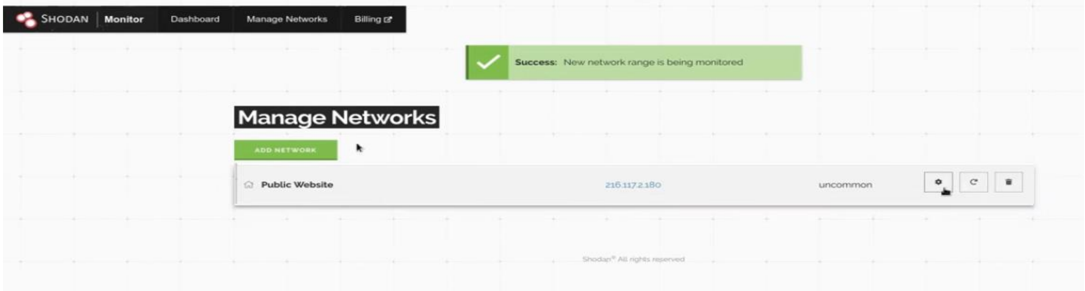
What is a trigger?
Triggers are rules that when they're met cause Shodan to send you a notification. For example, the 'malware' trigger will send you an email if the service looks like it has been compromised or it's running malware software.

For this site, we configure receiving alerts for unusual threats to services that should not be available to the public.


Notification Triggers

Select the types of notifications that you would like to receive. If none are selected we will let you know whenever Shodan discovers any service.

- □ **industrial_control_system** Services associated with industrial control systems
- □ **malware** Compromised or malware-related services
- ☑ **uncommon** Services that generally shouldn't be publicly available
- □ **open_database** Database service that does not require authentication
- □ **iot** Service associated with Internet of Things devices
- □ **internet_scanner** Device has been seen scanning the Internet and exposes a service
- □ **ssl_expired** Expired SSL certificate is used by the service



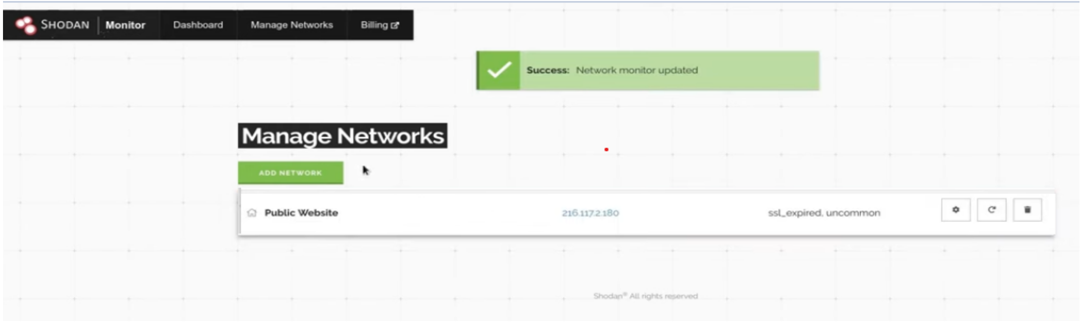
Success: New network range is being monitored

Manage Networks

Public Website	216.117.2.180	uncommon

Shodan® All rights reserved

At the same time, I want there to be notifications about the expiration of the site's security certificate (SSL certificate) and therefore I will modify the previous configurations with this new alert.



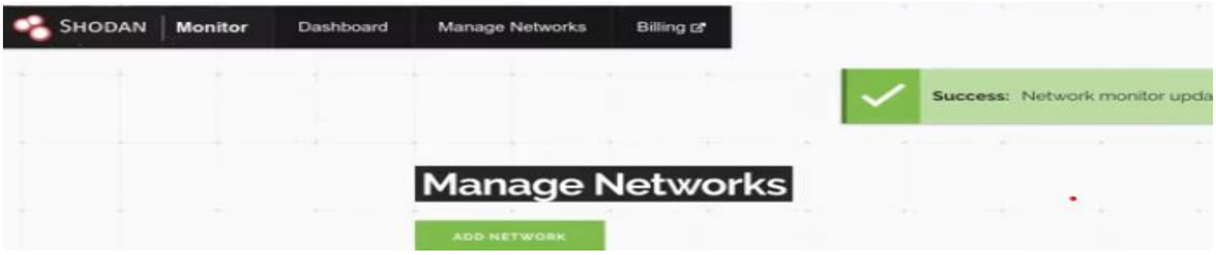
Success: Network monitor updated

Manage Networks

Public Website	216.117.2.180	ssl_expired, uncommon

Shodan® All rights reserved

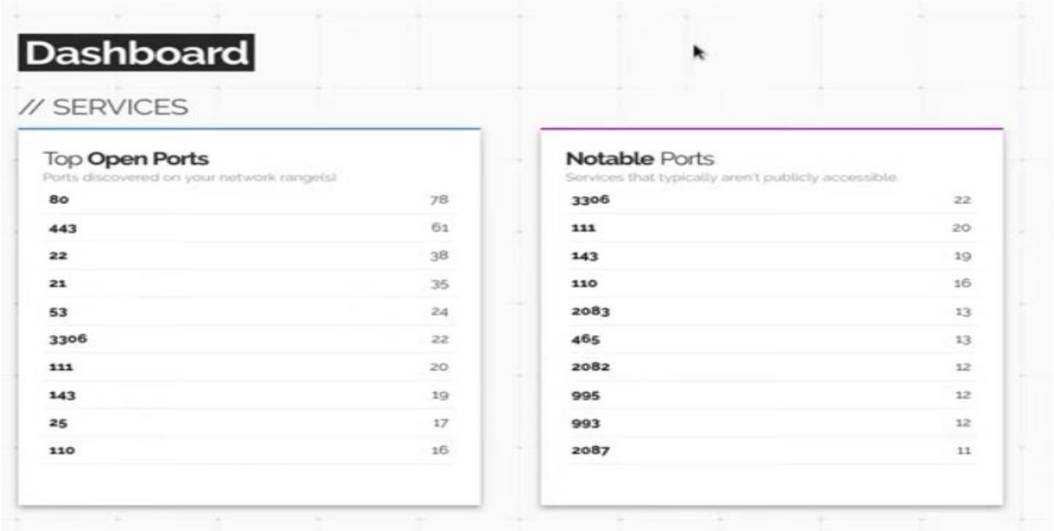
Once these events are configured, for which the alert is to be made, the possible alerts can be consulted in the Dashboard section, from the same interface.



Success: Network monitor upda

Manage Networks

In this section the information is structured as can be seen in the following image, and at first glance we see information about the open ports on the server.



Dashboard

// SERVICES

Top **Open Ports**
Ports discovered on your network range(s)

80	78
443	61
22	38
21	35
53	24
3306	22
111	20
143	19
25	17
110	16

Notable Ports
Services that typically aren't publicly accessible

3306	22
111	20
143	19
110	16
2083	13
465	13
2082	12
995	12
993	12
2087	11

3. Conclusion

Open-Source Intelligence (OSINT) is the practice of collecting information from published or otherwise publicly available sources. OSINT operations, whether practiced by IT security professionals, malicious hackers, or government-authorized intelligence agents, use advanced techniques to search through the flood of data available on the Internet and find that data in order to achieve objectives.

According to the OSINT Guide developed by the Romanian Intelligence Service [2] it is estimated that OSINT provides between 80% and 95% of the total data used by the intelligence community, worldwide [5].

OSINT provides access to some of the best data available in the world, whether you're conducting a research project, looking to gain competitive intelligence, uncover vulnerabilities, or conduct an analysis of potential threats.

Even if you are simply a person concerned about your privacy and want to find out what personal information has been inadvertently leaked, OSINT can be useful.

Open-source Intelligence (OSINT) can help organizations gather high-quality, grassroots intelligence and make choices based on it.

Open source, in this context, does not refer to the open-source software movement, although many OSINT tools are open source, instead, it describes the public nature of the data analyzed [7].

Despite their great utility, open-source intelligence tools also have a dark side that hackers or people involved in illegal activities can exploit, so great care and caution is required when using these tools, so as not to exceed legal limits.

References

[1]. https://www.govinfo.gov/content/pkg/USCODE-2011-title50/html/USCODE-2011-title50-chap15-subchapI-sec403-5.htm

[2]. https://www.sri.ro/upload/Ghid_OSINT.pdf

[3]. https://opensource.com/

[4]. https://www.imperva.com/learn/application-security/open-source-intelligence-osint/

[5]. https://sri.ro/

[6]. Truyens, Johan, Developing Open Source Capabilities, EDA Bulletin, nr. 9, iulie 2008

[7]. https://www.csoonline.com/article/3445357/what-is-osint-top-open-source-intelligence-tools.html

[8]. Clark, Robert M. 2013. Intelligence Collection. Sage Publications.

[9]. https://data.europa.eu/en/publications/datastories/open-source-intelligence

[10]. https://eda.europa.eu/

[11]. http://www.risk-uk.com

[12]. https://i-intelligence.eu/