Internal Security Review

2024, no. 31, pp. 383–412

https://doi.org/10.4467/20801335PBW.24.030.20807

ARTICLE

Open source intelligence on the internet – categorisation and evaluation of search tools

DANIEL MIDER

Faculty of Political Science and International Studies
University of Warsaw

https://orcid.org/0000-0003-2223-5997

Abstract

This article presents a comprehensive analysis and systematic review of search tools used in open source intelligence (OSINT). Three main categories of software were evaluated: systems integrated with operating system or web browser, standalone applications, and repositories of links to specialised tools. A critical evaluation of representative examples from each category was conducted, taking into account their functionality, effectiveness, and limitations. The analysis identified significant gaps in the current instrumentation and formulated postulates regarding potential directions for the development of the OSINT operator skills. The optimal development direction should focus on open source, modular tools with a low entry threshold, enabling community participation in their refinement and customisation for analysts' needs. The findings serve as a knowledge compendium for OSINT researchers, practitioners, and enthusiasts.

Keywords

open source intelligence, OSINT, cybersecurity, internet-based information retrieval

Introduction

There has been an unprecedented and continuous growth in the volume of data available on the internet, comprising both structured and unstructured information[1]. The growth in data has contributed to the widespread perception of the global web resources as a fundamental source of information and has led to the emergence and dynamic development of a methodology of information acquisition known as open source intelligence (OSINT) or white intelligence. This has resulted in the widespread use of OSINT in many fields: in cybersecurity, economic, competitive, political and military intelligence, criminal and insurance investigations, as well as in various forms of law enforcement support and risk management. The value and legitimacy of the use of open source evidence in international criminal investigations is increasingly emphasised[2]. Specialised, small-scale organisations using open source to document crimes and provide evidence with high degree of credibility play an important role in the collection of electronic evidence of crimes. For example, the findings of Bellingcat, a leading organisation dedicated to OSINT and investigative journalism, were deemed reliable enough to be used by the official commission investigating the downing of Malaysian airliner operating flight MH17[3]. In the ruling of the European

[1] The volume of data on the internet grew from 4.4 zettabytes in 2013 to 6.6 zettabytes in 2014. In 2020 it was 64.2 zettabytes and projections for 2025 indicate that it will reach 175-181 zettabytes. This means that the average annual growth rate is approx. 33%, with data volume doubling every 2.5 years or so. See: D. Reinsel, J. Grantz, J. Rydning, *The Digitization of the World. From Edge to Core*, https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf [accessed: 28 VI 2024]; *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025,* Statista, June 2021, https://www.statista.com/statistics/871513/worldwide-data-created/ [accessed: 28 VI 2024].

[2] A. Mackinnon, *Bellingcat Can Say What U.S. Intelligence Can't*, Foreign Policy, 17 XII 2020, https://foreignpolicy.com/2020/12/17/bellingcat-can-say-what-u-s-intelligence-cant/ [accessed: 28 VI 2024]; *Electronic evidence of war crimes. The role of journalists, media and social media*, webinar organised by Group of Friends on the Safety of Journalists and Media Freedom in Strasbourg and the Council of Europe, 25 XI 2022, https://www.coe.int/en/web/kyiv/-/electronic-evidence-of-war-crimes-and-the-role-of-journalists-media-and-social-media [accessed: 28 VI 2024]; E. White, *Closing cases with open-source: Facilitating the use of user-generated open-source evidence in international criminal investigations through the creation of a standing investigative mechanism*, Cambridge University Press, 7 IX 2023, https://www.cambridge.org/core/journals/leiden-journal-of-international-law/article/closing-cases-with-opensource-facilitating-the-use-of-usergenerated-opensource-evidence-in-international-criminal-investigations-through-the-creation-of-a-standing-investigative-mechanism/981CEFF9D5AF80B6FD0A75BE6A1A384C [accessed: 28 VI 2024].

[3] O. Matthews, *Fact Cats. The inside story of how it got the Skripal scoop,* The Spectator, 20 X 2018, https://www.spectator.co.uk/article/fact-cats/ [accessed: 28 VI 2024]. Also worth mentioning is the investigation of the April 2017 chemical attack in Khan Sheikhoun, Syria, which led to the death

Court of Human Rights in the case of Ukraine and the Netherlands v. Russia, the findings of the organisation running OSINT on topics such as: Russia's involvement in Ukraine, the fate of MH17, cross-border artillery attacks and the actions of Russian army personnel in eastern Ukraine were also among the evidence presented by lawyers representing Ukraine[4]. The effectiveness of OSINT methodologies has attracted attention of experts, and the information collected by civilian, independent organisations is highly regarded and used by intelligence professionals[5]. In the public discourse, there are opinions of special service officers that open source intelligence organisations are superior to the effectiveness of traditional intelligence services in many fields. However, the methodologies of these organisations are too innovative to be accepted by policymakers[6]. Experts further suggest that one organisation (Bellingcat) has become the Russian Federation's "biggest nightmare" as a result of the systematic disclosure of information about its internal and external activities[7]. The quoted statements have anecdotal status, but reflect the excitement of specialists aroused by the new phenomenon. The financial perspectives also look impressive – the global market for OSINT intelligence solutions is growing rapidly. In 2022, it was worth USD 4219.56 million and is forecast to reach USD 7317.89 million by 2031, which translates into a cumulative annual growth rate over the period

of around 100 deaths. Using open source data such as photographs, videos, meteorological data and eyewitness accounts, Bellingcat reconstructed the events before, during and after the attack. This approach identified the site of the rocket launch and incriminating evidence against Syrian government forces. The Bellingcat report was praised for the transparency of its methodology and the possibility of third-party verification of the evidence presented.

[4] E. Higgins, *How Open Source Evidence was Upheld in a Human Rights Court*, Bellingcat, 28 III 2023, https://www.bellingcat.com/resources/2023/03/28/how-open-source-evidence-was-upheld-in-a-human-rights-court/ [accessed: 28 VI 2024]; *Case of Ukraine and the Netherlands v. Russia, 8019/16, 43800/14, 28525/20*, Archive of the European Court of Human Rights, 30 XI 2022, https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-222889%22]} [accessed: 28 VI 2024].

[5] A. Mackinnon, *Bellingcat Can Say…* It is worth added that the analytical results provided by the OSINT community allow intelligence officers to discuss more freely in public without fear of revealing their sources or methods of obtaining information.

[6] In the original: 'I'd say they're way ahead of us on many things, (…) Bellingcat's methods are 'way too innovative for the great majority of lemmings in government (…).' See: O. Matthews, *Fact Cats…* Notabene, in 1996, the US Aspin-Brown commission recommended a change in the Intelligence Community's attitude towards OSINT, stressing that the US services have been slow to implement the methodology, and to date the postulated National Open Source-Intelligence Agency has not been established as the 19th member of the Intelligence Community.

[7] AFP, *How Bellingcat became Russia's 'biggest nightmare'*, France24, 7 IX 2022, https://www.france24.com/en/live-news/20220907-how-bellingcat-became-russia-s-biggest-nightmare [accessed: 28 VI 2024].

indicated of 6.31%[8]. The versatile use of open source intelligence in so many areas is due to its ability to efficiency acquire, process and analyse vast amounts of publicly available data, translating into valuable insights achieved at an unaffordable cost. The growing role of OSINT prompts an in-depth reflection on its relevance, potential and challenges of its application.

The purpose of this article is to categorise and evaluate selected OSINT tools for information search. The article focuses exclusively on open source software[9] solutions, as they are available for direct analysis by any user. This makes them central to OSINT research and practice. Open source tools should be the first choice in open source analysis due to their widespread availability and extensive community support. Creating a comprehensive map of subject areas in open source analysis provides an opportunity not only to systematise knowledge, but also to identify potential gaps in software and methodology. This process allows inferences to be made about possible gaps in these tools, especially in the context of meeting different information needs. Such an evaluation can identify areas for further development of adaptation.

The following research questions were formulated:

1. What are the gaps in the current OSINT open source toolkit and what are the most anticipated directions for development of these tools?
2. What are the functionalities and limitations of the various open source tools used in OSINT?
3. What open source tools for running OSINT are currently frequently referred to in training courses, instructional publications and typically found in author collections?

[8] *Open Source Intelligence Market Size, Share, Growth, and Industry Analysis, By Type (Video Analytics, Text Analytics, Visualization Tool, Cyber Security, Web Analysis, Social Media Analysis, and Others), By Application (Private Sector, Public Sector and Other), Regional Insights, and Forecast to 2032*, Business Research Insights, March 2024, https://www.businessresearchinsights.com/market-reports/open-source-intelligence-market-109546 [accessed: 28 VI 2024]. Cf. *Open Source Intelligence Market Size, Share, Competitive Landscape and Trend Analysis Report by Source, Technique and End User: Global Opportunity Analysis and Industry Forecast, 2020-2027*, Allied Market Research, May 2020.

[9] Issues relating to paid tools are beyond the scope of this article, as they are closed products that have the potential to put OSINT operators at risk. The nature of these tools as “black boxes” raises questions as to whether subject matter or search results are subject to registration by the provider, either explicitly or implicitly. Such risks can have serious implications for the security and confidentiality of analytical activities.

Definitional considerations

The term open source intelligence emerged in the practice of the US Intelligence Community, mainly the Central Intelligence Agency (CIA) and the Defence Intelligence Agency (DIA). In the civilian sphere, it only began to be used in the 1990s[10]. It was popularised at the time by Robert David Steele, a former CIA officer and author of *The Open-Source Everything Manifesto: Transparency, Truth and Trust*[11]. OSINT is translated using two equivalent terms – "open source intelligence" and "white intelligence". Although the OSINT method was used much earlier, there is still a lack of systematic historical studies devoted to it, and the available knowledge is often of anecdotal and sketchy character[12] and needs to be organised[13].

In line with the Aristotelian concept of *open source intelligence* will be defined by identifying its closest type (Latin *genus proximum*), which will allow it to be located in a broader context. The most commonly cited categorisation is normative, taking into account ethical and legal aspects. Thus, the term "white intelligence" is understood as a way of collecting data that raises neither ethical nor legal questions. According to available data, as much as 80% of the information acquired by intelligence services and data collection organisations today comes from open and unclassified sources. Its opposite – black intelligence, encompasses activities considered illegal in a given jurisdiction, often also unethical. The scope of black intelligence includes practices such as surveillance and infiltration using wiretaps, hacking, information and identity theft (including biometric), cryptographic security cracking, and

[10] A. Olcott, *Open Source Intelligence in a Networked World (Continuum Intelligence Studies)*, New York 2012, pp. 87–88.

[11] R.D. Steele, *The Open-Source Everything Manifesto: Transparency, Truth, and Trust*, Berkeley 2012.

[12] L. Block, *The long history of OSINT*, "Journal of Intelligence History" 2024, vol. 23, no. 2, pp. 95–109. https://doi.org/10.1080/16161262.2023.2224091; C. Colquhoun, *A Brief History of Open Source Intelligence*, Bellingcat, 14 VII 2016, https://www.bellingcat.com/resources/articles/2016/07/14/a-brief-history-of-open-source-intelligence/ [accessed: 28 VI 2024].

[13] The value of open sources was recognised in the United States as early as the 18th century during the American Revolution. George Washington drew information about the strength of British troops and their activities from newspaper publications. In turn, during World War II, General George Patton's troops used maps found in Michelin filling stations. The British government in 1939 asked the BBC media corporation to set up a commercial summary service of foreign press and radio media called Digest of Foreign Broadcasts (now BBC Monitoring). After World War II, think-tanks were widely set up and institutions were created to obtain information from open sources and analyse it (this was done by the East German Stasi, the Chinese, who set up the Institute of Scientific and Technical Information of China, and the Americans with Sherman Kent, the historian known as the father of the American intelligence school). Cf. F. Schaurer, J. Störger, *Guide to the Study of Intelligence. The Evolution of Open Source Intelligence (OSINT)*, "The Intelligencer: Journal of U.S. Intelligence Studies" 2013, vol. 19, no. 3, p. 53.

information acquisition using blackmail and corruption, among others. It is estimated that around 5% of political and economic intelligence is obtained through these methods[14]. Situated between white and black intelligence is the category of grey intelligence, encompassing activities that cannot be clearly classified in terms of legality and ethics. Although, by definition, these methods are legal, they are contrary to ethical principles, whether universal or specific. Exemplification of grey intelligence can include surveillance in the form of observation and monitoring of persons or vehicles, as well as sociotechnical activities aimed at manipulation and phishing. Estimates suggest that around 15% of intelligence is obtained through grey intelligence methods. The boundary between legality and illegality, as well as between ethical and unethical, is sometimes blurry and dependent on the cultural, social and political context. Several definitions of open source/white intelligence are widespread in the academic, intelligence and journalistic circuits. The most commonly cited is the CIA's ostensive definition, indicating that OSINT is a type of intelligence based on sources that are legally available to a large number of people and include their wide repertoire: media, literature, first, second and third sector reports, satellite photographs, maps, geodetic documents, geographical and meteorological data, as well as scientific, technological, industrial and social and demographic information[15]. This definition is accepted as a working definition, with the knowledge that other interpretations exist[16].

From a historical and technological perspective, three cumulative, chronologically positioned stages in the development of open source intelligence can be distinguished: OSINT 1.0, OSINT 2.0 and OSINT 3.0. The first two co-exist

[14] The percentage breakdown indicated, referring to the share of open source intelligence in the political and economic intelligence structure, is partly anecdotal. It is based mainly on estimates presented in a publication by Arthur S. Hulnick, a former CIA officer and lecturer at Boston University. See: A.S. Hulnick, *Fixing the Spy Machine. Preparing American Intelligence for the Twenty-First Century*, Westport 1999, pp. 40–41. Contemporary, relevant publications either replicate these proportions (e.g. S.C. Mercado, *Sailing the sea of OSINT in the information age*, "Studies in Intelligence" 2004, vol. 48, no. 3, pp. 45–55), or point vaguely to the growing role of open-source intelligence (e.g. R.D. Steele, *Open source intelligence*, in: *Handbook of Intelligence Studies*, New York 2007, pp. 129–147).

[15] *A Consumer's Guide to Intelligence*, Office of Public Affairs CIA, 1999, https://archive.org/details/consumersguide_tenet/mode/2up [accessed: 28 VI 2024]. The definition adopted deliberately ignores issues of overlap and intersection of other types of intelligence, e.g. IMINT, HUMINT, GEOINT, with open source intelligence so understood.

[16] The following are also apt definitions of white intelligence: NATO (*NATO Open Source Intelligence Handbook v 1.2*, https://archive.org/details/NATOOSINTHandbookV1.2/page/n1/mode/2up [accessed: 28 VI 2024]) and Mark M. Lowenthal's (M.M. Lowenthal, *Intelligence: From Secrets to Policy*, Washington 2007).

side-by-side, while the third is partly predictive in nature. In this article, the author considers them in descriptive terms of the evolution and development of OSINT.

OSINT 1.0 – encompasses the early 1980s, when intelligence analysts personally searched and "manually" aggregated sources of information (primarily "analogue", less often digitised), conducting searches and analyses with little assistance from information technology (directories and databases and the first pre-WWW internet networks, such as USENET, BBS or FTP, and the early WWW).

OSINT 2.0 – associated with the development of the internet, primarily social media (the so-called Web 2.0 and Web 2.5). In OSINT 2.0, analysts use search engines, news aggregators, social network analysis and process automation[17]. In this period – currently ongoing – not only has there been an increase in the amount of information available, but an overabundance of information has appeared[18].

OSINT 3.0 – is the latest, partially anticipated phase of OSINT, which uses analytical tools from big data, machine learning and natural language processing, as well as weak artificial intelligence. OSINT 3.0 is dominated by advanced algorithms for searching and analysing in real time from huge amounts of data and drawing conclusions. In addition, the area of Web3 and metaverse is being explored, and there is a convergence of OSINT and other types of intelligence, including geospatial intelligence (GEOINT), human intelligence (HUMINT), measurement and signature intelligence (MASSINT), signals intelligence (SIGINT). It is possible that one future trend will be the popularisation of decentralised, bottom-up team structures of OSINT researchers, operating on the principle of crowdsourcing. This is an area *in statu nascendi*[19].

[17] H.J. Williams, I. Blum, *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*, RAND, 17 V 2018, https://www.rand.org/pubs/research_reports/RR1964.html [accessed: 28 VI 2024].

[18] Cf. K. Tylutki, *The information of a mass destruction range – OSINT in intelligence activities,* "Internal Security Review" 2018, no. 19, pp. 384–404.

[19] A.W. Dorn, *United Nations Peacekeeping Intelligence*, in: *The Oxford Handbook of National Security Intelligence*, L.K. Johnson (ed.), Oxford 2010, p. 280; D. Mider, J. Garlicki, W. Mincewicz, *The Internet Data Collection with the Google Hacking Tool – White, Grey or Black Open-Source Intelligence*, "Internal Security Review" 2019, no. 20, pp. 280–300; K. Turaliński, *Wywiad gospodarczy i polityczny. Podręcznik dla specjalistów ds. bezpieczeństwa, detektywów i doradców gospodarczych* (Eng. Economic and political intelligence. A handbook for security professionals, investigators and economic advisers), Warszawa 2015, pp. 31–33.

Open source intelligence tools

Classifying current open source intelligence software is challenging due to the heterogeneity of functions, diversity of methodologies and the dynamic development. The number and type of possible ways to explore the internet are limited, so that the functions of the various programmes overlap. The number of programmes and other OSINT tools is estimated by the author to be approx. 1000-1200. The main problem is the rapid obsolescence of the tools – for reasons both internal (loss of motivation by team or its individual members) and external (blocking specific channels for information acquisition). Essentially, three groups of OSINT tools can be distinguished: basic, i.e. search tools; support tools, i.e. analytical and visualisation tools; as well as security tools, which provide the OSINT operator with anonymity or a false but consistent identity.

OSINT search tools include most of the tools and methods used in open source intelligence. Many experts identify OSINT primarily with these tools, as they enable the effective collection of information from open sources. These tools can be divided into three main categories: those integrated into the operating system or browser, standalone applications that require registration and configuration, and links to various OSINT resources that the user must install and configure themselves.

OSINT analytical and visualisation tools play the most important role in processing, analysing and visualising data collected from open sources. They help to identify patterns and trends, and enable the creation of charts, maps and other forms of visualisation to facilitate data interpretation.

OSINT security tools provide protection when collecting and analysing data from open sources. These include privacy enhancing technologies (PET) and tools to detect and avoid risks associated with the processing of publicly available information. The software can be categorised from the perspective of the ergonomics of the OSINT tool operator.

The author distinguished three groups of search tools giving the operator qualitatively different experiences (learning curve, efficiency, flexibility) and posing various barriers (usage and cognitive).

1. **Software integrated into the operating system and/or web browser**. The OSINT operator receives a finished, secure product containing numerous tools. Most often, these are already pre-installed and the initial configuration enables them to be used[20].

[20] A review of existing solutions shows that the choice of underlying platforms for integrated OSINT tools is carefully considered and security issues are addressed with detail.

2. **Standalone software, installed locally or online** (server of the provider). Before using such a tool, the operator usually has to register the installation, obtain API (application programming interface) keys and configure the tools. Additional software such as SQL (structured query language) is sometimes required.
3. **Links to software and other OSINT tools.** These can function as tabs, worksheets, webpages, other types of lists. The OSINT analyst is provided with a list of classified and sorted tools. He or she is provided only with links. Any installation, configuration or other necessary steps related to the use of the service must be done by himself.

Representative programmes within each of the three groups were analysed.

OSINT software integrated with an operating system and/or web browser

This is the least numerous group of software (and therefore discussed at length), and requires the greatest input from the supplier and relatively the least from the OSINT operator. This is a collection of solutions currently being developed. The systems discussed here include tools more broadly described in later sections of the article.

Trace Labs (Crowdsourced Open Source Intelligence for Missing Persons) – is the successor to Linux Buscador[21]. Trace Labs (TL) was founded in Canada in 2017 by Adrian Korn, security consultant, auditor, author of monographs in the OSINT area. It is a non-profit organisation dedicated to collecting information on missing persons, and Korn is currently its CEO. Trace Labs is one of the world's leading OSINT projects and greatly enhances search efficiency. Through the implementation of crowdsourcing, it engages a global community of volunteers to work with law enforcement and missing persons organisations. The aim is to support official search efforts, not replace them. Trace Labs OSINT is derived from Kali Linux distribution, and this one – from Debian. The system runs using a virtual machine, which is part of the security assurance, and is given in the universal (Oracle VB, VM Ware) OVA (open virtualisation appliance) format[22]. Kali Linux base system has been significantly

[21] Trace Labs once developed the Linux Buscador operating system, used by many government agencies, private companies and NGOs. Buscador included numerous tools related to digital intelligence, network penetration and social media monitoring. However, it ceased to be developed in favour of Trace Labs OSINT. No official information was given on the reasons for the cancellation. In a blog post, Adrian Korn indicated that the decision was driven by the need for a more advanced distribution that would better suit the needs of the organisation.

[22] The project's homepage: https://www.tracelabs.org/initiatives/osint-vm. Download: tlosint-vm, in: https://github.com/tracelabs/tlosint-vm/releases [accessed: 28 VI 2024]. There is no direct way to run the system in the most secure environmenti.e. Qemu/KVM. It is necessary to convert the .ova file to .qcow2 or compile from source. In addition, the system is not ready for use immediately after start-up, the tools have to be installed using a desktop installation script (install-tools.sh). Only then is full functionality achieved.

modified by the introduction of numerous limitations in the original product – of the wide repertoire of programmes, only a few have been left. However, Trace Labs OSINT has other software such as CherryTree, EyeWitness, ExifTool, Maltego, Metagoofil, OSRFramework, Recon-ng, Spiderfoot, TheHarvester, TcpDump, Wireshark. In addition, two web browsers, Chromium and Firefox ESR, stocked with several hundred links for open source intelligence, were included in the project[23]. A strength of the collection of these links is the wide range of software used to search for people (the main target of the TL) and telephone numbers. The large (but not exhaustive) number of search engines, clearly grouped, is noteworthy. A weakness is the paucity of cartographic references and those related to social networks outside the mainstream, e.g. the absence of Fediverse, an alternative social media system. Furthermore, there are no references to overlay networks (e.g. The Onion Router – Tor, Invisible Internet Project – I2P, Lokinet), alternative DNS systems (e.g. OpenNIC) and underground forums (imageboards) and chat rooms. For the novice user, the need to install tools and libraries using a terminal will be an obstacle to using the system. For the advanced user, on the other hand, a major limitation may be the inability to run the system directly under the control of the most secure hypervisor, which is Qemu/KVM. However, the usefulness of this tool and the fact that it is the main one of the comprehensive OSINT tools should be emphasised. For these reasons, it has been discussed at such length. The shortcomings identified appear to be closely related to the profiling of the system to meet the specific operational requirements arising from the exploration mission. This justifies certain compromises in terms of installation and security.

Tsurugi Linux[24] – is a GNU/Linux operating system distribution designed for DFIR (digital forensics and incident response) and malware analysis activities developed in 2019 by a team of developers and security professionals from Japan, the US, UK, Germany and Australia. The team leader is Mati Aharoni, previously working on the Kali Linux (Backtrack) operating system, founder and CEO of Offensive Security. The company is based in the US and is funded by the sale of cyber security tools and undisclosed private donations. The system does not focus on OSINT software, but includes tools related to network security, computer

[23] They were categorised as follows: Company, Internet Scan, Email Search, Phone Number, People, Maps & Geography, Search, Social Media Tools, Social Networks, User Name Check, Collections, Broad Search Tools. In addition, two practical plugins have been added to Firefox - the first allows websites to be read aloud in more than 40 languages, while the second removes tracking elements from web links. In addition, several other plugins have been introduced to personalise the browser.

[24] Tsurugi Linux, https://tsurugi-linux.org/index.php [accessed: 28 VI 2024]. Tsurugi (剣) is a Japanese word meaning ‘sword’.

forensics and data recovery[25]. It can successfully compete with the well-known and widely used Kali Linux in this area[26]. Tsurugi Linux is based on the Ubuntu distribution (of the Long Term Support variety), meaning that it uses its stable, long-term supported components. The system's kernel includes optimisations to support recovery tools and malware analysis. There are (usually) two releases of the system each year in the development cycle. The tool is offered in different formats – in one cycle as an ISO image, in another as an .ova file. This results in inconsistency in the format of the releases made available. The system is designed to run both in a virtual environment (on a virtual machine) and on physical hardware, as it can be installed directly on the hard drive. In both cases, installation is straightforward and the tool is immediately operational. It is clearly a dual-use technology, as it can be used for both legal and illegal activities[27]. Among the 16 groups of tools offered by Tsurugi Linux, one is dedicated to open source intelligence. The content is dozens of selected elements – individual tools, aggregators and frameworks. The vast majority of the collection are those that operate on the command line. A valuable feature is the desktop-based OSINTSwitcher, which allows switching between a search order for cyber security and a search order for OSINT, making the operator's job easier. In Tsurugi Linux, the basis for the work of OSINT is a modified Mozilla Firefox browser called OSINTBrowser. The browser is equipped with links to 11 groups of OSINT tools[28], and approx. 30 plugins (extensions) have been added for purposes ranging from security to OSINT. Most extensive are the lists of tools for geographical location (aerial maps, satellite maps and other geo-location methods) and air, sea and rail transport tracking. Valuable additions are a collection of links to OSINT analytical tools and a set of information aggregation tabs on radio frequency

[25] It should be added that alongside Tsurugi Linux, software such as the Bento DFIR Portable Toolkit and Tsurugi Acquire are also offered. Bento DFIR is a portable toolkit designed for use in investigative analysis and security incident response. This kit contains a wide range of tools necessary to carry out various operations in the field of digital forensic analysis for intuitive and convenient use even by an intermediate user. Tsurugi Acquire, on the other hand, is designed for digital forensic analysis and data recovery. It is a lightweight, 32-bit version of the system that allows you to easily perform data retrieval operations from various storage media. All tools are offered free of charge.

[26] The advantage of Tsurugi Linux over Kali Linux is that Tsurugi is equipped with an advanced forensic analysis module, including cloud-based artefacts, photographs, system images and other digital artefacts. In addition, it offers a comprehensive set of tools for forensic analysis of cryptocurrencies.

[27] Cf. National Research Council, *Computers at Risk: Safe Computing in the Information Age*, Washington 1991; J. Forge, *A Note on the Definition of "Dual Use"*, "Science and Engineering Ethics" 2010, vol. 16 no. 1, pp. 111–118. It appears that the developers may therefore have decided to introduce elements into the system that could potentially de-anonymise users.

[28] They were categorised as follows: Resources, Geo Based Searches, Metadata, Socials, P2P, Transport, Date-Time, Website analysis, Search engine, Radio, Commercial Registries.

scanning/monitoring and radio broadcasting[29]. A weakness is the small number of search engines (there are only a dozen or so). The choice of the Firefox browser limited the possibilities of using the many specialised plugins that are available in more abundant Blink-based family (Chrome, Chromium, Brave, MS Edge). The implementation of bookmarklets[30] and proprietary plugins was also abandoned. Tsurugi Linux is also deficient in the following areas of analysis: overlay networks (it has an extremely small number of Tor exploration tools), alternative social media and alternative domain systems. However, its uniqueness in terms of specific application possibilities should be highlighted.

Many other operating systems for penetration testing also have modules containing tools designed for OSINT. These are dual-use tools[31].

BlackArch Linux[32] – is an operating system initiated in 2013 by a team of volunteer cyber security enthusiasts. The project is mainly funded by voluntary donations from users and the community to cover hosting, domain name maintenance, mirrors and test hardware. BlackArch is based on the Arch Linux distribution, thus utilising the principles of the KISS (Keep it Simple, Stupid) rule, the essence of which is to achieve the highest performance in combination with ease of use. The distribution is available in both a full ISO version, offering a variety of window managers, and a "Slim" version with the XFCE desktop environment. The system is recommended for intermediate users familiar with Arch Linux. BlackArch is a comprehensive penetration testing and cyber security research solution with an impressive number of nearly 3,000 tools focused on IT security, including exploitation, forensic analysis, penetration testing, reverse engineering, network analysis and more[33]. Among the wide range of tools are also those aimed at OSINT tasks, but a separate category has not been created for them[34].

[29] These include, for example: SDR.hu - a web portal that provides access to remote SDR (software defined radio) radio receivers worldwide. Users can use a browser to listen to and control SDR receivers from different locations across a wide range of radio frequencies. Broadcastify - an online platform for listening to live radio broadcasts.

[30] Bookmarklet is a small, lightweight JavaScript script saved as a web browser tab. It allows you to perform tasks on the current webpage without having to install plugins or extensions.

[31] It is worth noting that 2013 was a turning point for IT security software due to the need to upgrade existing tools, the growing importance of cyber security (vide: the Edward Snowden case) and the growth of the open source community.

[32] BlackArch, https://blackarch.org/index.html [accessed: 28 VI 2024].

[33] Toolkit available at the following address: https://blackarch.org/tools.html.

[34] These tools have been categorised according to their purpose. They include, but are not limited to: exploitation (186 tools), scanners (313 tools), web applications (310 tools), password cracking (169 tools) and investigative analysis (129 tools).

ParrotOS Security[35] – was publicly released on 10 April 2013 by Lorenzo Faltra, team leader and core developer. The system originated from the community forum Frozenbox, also created by Faltra. ParrotSec is a community interest company[36] registered in the UK, based in Palermo, Italy. ParrotOS Security is *open source* system, funded mainly by community and volunteer donations. Based on Debian, *the testing* branch, it uses a rolling release model. It is a distribution for penetration testing, security research, forensic analysis, reverse engineering and cryptography. The system can run on servers, desktops, laptops, virtual machines and IoT (Internet of Things) devices, including the Raspberry Pi. It offers over 600 tools, including full disk encryption and privacy protection (Tor and AnonSurf). An advantage is the forensic mode, which prevents automatic installation of storage devices, protecting sensitive data from modification. It is used by government agencies and law enforcement agencies. It has a modest OSINT menu resource in a separate tab, with programmes such as Censys, cloud-enum, emailharvester, inspy, installooader – Instagram OSINT tool, Maltego, sherlock, Shodan and TheHarvester[37].

Kali Linux[38] – made its market debut in March 2013 as the successor to BackTrack Linux. The development and maintenance of this distribution is the responsibility of Offensive Security, which profits from providing certified training and courses, such as the renowned Offensive Security Certified Professional. As a comprehensive tool for security professionals, Kali Linux includes over 600 pre-installed tools. These include a modest OSINT Analysis tab, which includes tools such as Maltego, SpiderFoot and TheHarvester. Kali Linux runs on numerous platforms, including virtual machines. It is characterised by a high frequency of updates, precision and design aesthetics.

In cybersecurity-focused Linux distributions such as BlackArch, Kali and Parrot, OSINT tools are an integral part of the arsenal, although their number and prominence do not match the other categories. The exception is Kali Linux, which stands out from the other systems by offering a tab dedicated to OSINT tools. By analysing the tools in the distributions described, it can be seen that the universe of cyber security and OSINT intermingle, sharing common instruments.

[35] ParrotOS Security, https://www.parrotsec.org [accessed: 28 VI 2024].

[36] One of the legal forms of UK social enterprises (editor's note).

[37] Worth mentioning are cloud-enum, a tool for enumerating and analysing cloud resources such as S3 buckets or EC2 instances, and tools for analysing accounts on Instagram: inspy and instaloader.

[38] Kali Linux, https://www.kali.org [accessed: 28 VI 2024]. The name comes from the Hindu goddess Kali, commonly associated with strength and power.

CAT (Cyberuniverse Analysis Tool)[39] is a Polish collection of OSINT tools functionally integrated into the Brave browser[40]. CAT was created because there was a need for a tool that was lightweight, quick to install and use, capable of running in any environment and on any machine, partially anonymised and not indicative of service activity. Work on it began in June 2021. By 2024, a version had not yet been produced to the full satisfaction of the developers. The tool is used in detective investigations, KYC checks and academic searches, teaching and OSINT training. CAT is made available under a ♡Copyheart license[41]. The following goals guided the creation of the tool: gentle learning curve (ease of use after a short training), speed of installation (installation in a quarter of an hour), convenience (integration of all tools in a single browser), versatility (versatile use in various investigations and research), jurisdiction (Polish; no similar solutions currently available), scalability (simultaneous operation of multiple users, versioning), flexibility (compatibility with different systems, limited change of digital fingerprint), reconfigurability (customisation of tools, change of functionality and appearance), modularity (easy addition and removal of components), open source and free of charge (free, open source tools, local operation), consideration of socio-cultural and linguistic aspects (exploration of subcultures, sociolect dictionaries), relative anonymity (approx. 1,500-2,000 users), anonymity in the team (collaboration without revealing identity). Each of these objectives has been at least partially met. The tool is not, however, free from defects. Firstly, not all plugins and links are audited, creating a real, albeit low, risk of surveillance (the Brave browser, considered one of the most secure, improves security considerably). Secondly, the use of mainly open source and free tools may restrict access to some data. Thirdly, CAT is currently developed by a single person and with own funds.

The following tools, services and solutions have been implemented into CAT: over 700 online venues and tools, over 130 plugins, over 20 bookmarklets[42], over 270 search engines, over 100 Deep Web & Darknet venues, over 60 Polish databases/search engines/venues, over 50 security/anonymisation (going grey) tools, over 40 social media stalking/doxing tools, over 30 Imagery/Map Intelligence tools, over 30 Crypto/Blockchain Intelligence tools, over 20 educational tools, over 20 fake news verification tools, over 10 leak sites, over 10 alternative social media, 8 OSINT

[39] Tool available upon request via email: d.mider@uw.edu.pl.

[40] It refers conceptually to the Oryon OSINT browser created by Polish infobroker Marcin Meller. This tool ceased to be updated in 2017. See: https://sourceforge.net/projects/oryon-osint-browser/ (attention: version of 6 IV 2017).

[41] N. Paley, *Copying is an Act of Love*, https://copyheart.org [accessed: 28 VI 2024].

[42] Systematically replaced by proprietary plugins.

frameworks. In addition: hacker forums, data processing, data analysis, warez sites, *privacy enhancing technologies*, *operations security* (OpSec) tools; automatic support for Tor networks, IPFS (InterPlanetary File System) with Unstoppable Domains; scripts for hidden Polish imageboards (Karachan and Wilchan, including links to hidden "elite" boards); complete access data for Hyphanet, Lokinet and Alternative Top Level Domains (ATLD) overlay networks.

Several elements distinguish the CAT compared to existing OSINT solutions. Firstly, it was considered necessary to implement a complete set of sources concerning Poland. Most of the resources of the OSINT community are characterised by global or Euro-Atlantic reach. Countries experiencing geopolitical tensions or taking controversial actions are also subject to intensive OSINT analysis. However, the global OSINT community's knowledge of Central and Eastern Europe, particularly Poland, is limited. There are numerous resources for white economic and personal intelligence that are unnoticed in the world. Secondly, Deep Web and Darknet[43] tend to be on the borderline of interest for OSINT operators. They are generally associated (in the public's mind and by publicists) with illegal activities, including illegal information gathering, which is partly a misconception. A second problem is that OSINT analysts limit themselves to only one network, namely Tor, while several more, albeit smaller ones, are in operation[44]. Several scholars emphasise, admittedly, the importance of the Deep Web and the Darknet for OSINT, but these are not widely known voices. Moreover, analysts treat these phenomena narrowly, limiting OSINT exploration to the acquisition of information about criminals and crimes in order to combat them[45]. A number of search tools in this area have been implemented in CAT. Thirdly, science intelligence (SCINT) plays an important role in both the state and commercial intelligence sectors[46]. It is mainly based on open sources, encompassing

[43] These terms are explained in: D. Mider, *Mappa Mundi ukrytego Internetu. Próba kategoryzacji kanałów komunikacji i treści* (Eng. Mappa mundi of the hidden internet. Categorising internet communication channels), "PTINT Praktyka i Teoria Informacji Naukowej i Technicznej" 2015, vol. 23, no. 1, pp. 3–16.

[44] This misunderstanding is attempted to be clarified by the authors: V. Ciancaglini et al., *Deep Web and Cybercrime: It's Not All About TOR*, https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/deep-web-and-cybercrime-its-not-all-about-tor [accessed: 28 VI 2024].

[45] Cf. M. Bazzell, *Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information*, CreateSpace Independent Publishing Platform, 2018; M. Chertoff, T. Simon, *The Impact of the Dark Web on Internet Governance and Cyber Security*, https://www.cigionline.org/static/documents/gcig_paper_no6.pdf [accessed: 28 VI 2024].

[46] P. Maddrell, *Spying on Science: Western Intelligence in Divided Germany 1945–1961*, Oxford 2006; *Wyniki pracy wywiadu naukowo-technicznego MSW PRL 1971–1989* (Eng. The results of the operations of the scientific-technical intelligence of the Polish People's Republic 1871-1989), M. Sikora (comp.), Katowice–Warszawa 2019; H. Nasheri, *Economic Espionage and Industrial Spying*, Cambridge 2004; K. Turaliński, *Wywiad gospodarczy i polityczny*…

the process of acquiring and analysing scientific information to optimise decision-making processes, innovation strategies and science policy. Of particular relevance here is grey literature, defined as information materials and publications not covered by traditional commercial publishing channels, which provide unique data and conclusions despite difficulties in indexing and access. Understanding the value of these resources, CAT integrates both traditional scholarly sources and grey literature into its activities. Fourthly, search engines are the most common, yet most misused, data search tool[47]. The Google search engine can be considered in terms of the most easily used hacking tool[48]. The introduction of this area of OSINT stems directly from the framing of the internet as a reservoir of data[49]. It should also be noted that the most popular search engine, Google, is not so much used as abused by users who overlook other global, local or specialised search engines[50]. Such search engines are numerous, covering a different area than Google does. They are included in the CAT, including search engines and tools that use large language models (LLM). Fifth, leaks of government and corporate documents are an important area of OSINT analysis. The rapid identification, access and evaluation of these materials are critical for the institutions concerned, political actors and societies as a whole. The phenomenon of whistleblowers, although already present in the pre-internet era, has gained importance with the development of digital technologies[51]. Examples such as the actions of Daniel Ellsberg, Chelsea Manning or Edward Snowden demonstrate the scale and impact of this phenomenon on contemporary geopolitics and public discourse. The internet has enabled the emergence of specialised platforms and channels for the distribution of information leaks. For OSINT operators, these sources are a critical element in the data acquisition and analysis process and are

[47] This problem was recognised earlier. See: E. Hargittai, A. Hinnant, *Digital Inequality: Differences in Young Adults' Use of the Internet*, "Communication Research" 2008, vol. 35, no. 5, pp. 602–621. https://doi.org/10.1177/0093650208321782). It still appears to be up to date. See: K. Abramczuk, M. Kąkol, A. Wierzbicki, *How to Support the Lay Users Evaluations of Medical Information on the Web?*, in: *Human Interface and the Management of Information: Information, Design and Interaction*, S. Yamamoto (ed.), Cham 2016, pp. 3–13. https://doi.org/10.1007/978-3-319-40349-6_1.

[48] Cf. D. Mider, *Sztuka wyszukiwania w Internecie – autorski przegląd wybranych technik i narzędzi* (Eng. The art of searching on the internet. Review of selected techniques and tools), "Studia Politologiczne" 2019, vol. 54, pp. 191–229; D. Mider, J. Garlicki, W. Mincewicz, *Pozyskiwanie informacji z Internetu metodą Google Hacking…*

[49] M. Bazzell, *OSINT Techniques: Resources For Uncovering Online Information*, [n.p.] 2023.

[50] *Market share of leading desktop search engines worldwide from January 2015 to January 2024*, Statista, 2024, https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed: 28 VI 2024].

[51] P. Rosenzweig, T.J. McNulty, E. Shearer, *Whistleblowers, Leaks, and the Media: The First Amendment and National Security*, Chicago 2015.

therefore included in the CAT. As part of a comprehensive approach to open-source analysis, additional subject areas are included. The scope includes analysis of cryptocurrencies and blockchains in the context of OSINT, as well as open sources for economic intelligence, including corporate intelligence, financial intelligence and trade intelligence. In addition, cartographic sources relevant to OSINT and search engines for human sexuality content are included. The integration of these diverse research domains enables a multi-faceted approach to open source analysis.

A few words should be added about the choice of carrier, i.e. the Brave browser. After analysing web browsers, Brave was chosen as the optimal one. It stands out for its high level of privacy and security, as confirmed by the author's tests conducted with BrowserAudit[52] and PrivacyTests[53]. A significant strength of Brave is its open source nature, and the browser's security is further strengthened by the *bug bounty* programme. The browser is characterised by its versatility – it is compatible with various operating systems. The use of the Blink engine allows access to a wide range of extensions, which increases the functionality of the tool. A team of experienced specialists, led by Brendan Eich, creator of the JavaScript language, is responsible for the development of Brave, which is a security recommendation. In addition, Brave Software bases its business model on an innovative BAT (Basic Attention Token) token system. This eliminates the need to monetise user data.

Standalone OSINT software, installed locally or online

This is the group of OSINT software that is relatively the most numerous and has the broadest applications. It is difficult to determine who started the habit of sharing resources, as there are no systematic studies on the history of the global open source intelligence community. However, one can point to Arno Reuser. He is an "intelligence librarian" and one of the pioneers of OSINT. His work dates back to the 1990s. He created an intelligence library on Reuser's Information Services (RIS) website containing OSINT resources. The website is no longer available and RIS has evolved into a consultancy. Undoubtedly, mention should also be made of the well-known OSINT pioneer and populariser Michael Bazzell. He worked for nearly two decades in various US federal and local law enforcement units. His resources are widely recognised as among the first in the field. Bazzell is the author of a number of books on OSINT and counter surveillance on the internet. He hosts the popular podcast *The Privacy, Security, & OSINT Show*, is the creator of IntelTechniques[54],

[52] BrowserAudit, https://browseraudit.com [accessed: 28 VI 2024].

[53] PrivacyTests, https://privacytests.org [accessed: 28 VI 2024].

[54] IntelTechniques, https://inteltechniques.com [accessed: 28 VI 2024].

a platform with tools and training for OSINT professionals. The best, most used and therefore representative software was analysed.

Maltego[55] – was developed by Paterva, founded in 2007 or 2008 (various sources state differently) in South Africa by Roelof Temmingh, an expert in cyber security, information intelligence and data analysis. It is one of the most popular and respected OSINT tools used by government and military agencies[56] as well as private companies around the world[57]. The emphasis is on a user interface that is clear and gives an in-depth overview of the information. Maltego is one of the most convenient, intuitive and at the same time aesthetically pleasing tools. It is integrated, installed on the computer's hard drive and used either from the producer's server (free version and cheaper paid versions[58]) or from its own server (in which case the search topics are invisible to the producer). Maltego has evolved from a simple business relationship mapping tool to a sophisticated OSINT software. It is not open source, but has been discussed because of its widespread use. The foundation of its operation is a triad of closely related elements: entities, collections and transforms, the synergy of which enables efficient retrieval and processing of data. Entities, which are elementary units of information (e.g. person, website, IP address), form the basic building blocks of the system. Collections, which are sets of correlated entities (e.g. company employees, domain email accounts), introduce the structural dimension. On the other hand, transformations, which are automated operations that process entities (e.g. geolocalisation based on IP address, WHOIS data extraction), are responsible for the flow and enrichment of information. The result of the interaction of these components is a clear graph of relationships, offering a holistic perspective. A separate aspect of the tool's functionality are transformations not integrated into Maltego (plugins). These are extensions that can be quickly and easily launched to enable users to perform analytical tasks immediately. The extensive collection of plugins (currently just over 80), a list of which is presented on the home screen of the application, is available on a variety of terms and conditions – ranging from click-and-run unlimited use, to free options, albeit requiring an API key, to subscription options involving fees. Through plugins, Maltego integrates many well-known tools serving OSINT[59]. Among the plugins with unrestricted access,

[55] Maltego, https://www.maltego.com [accessed: 28 VI 2024].

[56] These include: Europol, FBI, NATO, police in the UK, Scotland Yard.

[57] E.g. consulting and auditing companies such as: Deloitte, EY, KPMG, PwC and dealing with cyber security: Mandiant, Palantir, Recorded Future.

[58] Including Maltego Government/Military for governmental institutions.

[59] Among the many are blockchain explorers, email address search software – Hunter, or online aircraft and marine tracking systems.

one that catches the eye is Maltego Regex Transform (Maltego Technologies), which enables the extraction of matching objects from websites using regular expressions. Another plugin is Loginsoft OSINT, which detects phone numbers and extracts metadata. Also present is a group of plugins to help identify links and relationships. LittleSis (Maltego Technologies) enables the identification of relationships between individuals in the business and government spheres (mainly US). OCCRP Aleph (Maltego Technologies) allows searches of business records, documents, public procurement data, sanctions, leaks, news articles and other resources, providing one of the best sources for investigative journalists. Social Links CE (Social Links Inc.) provides the ability to extract data from IoT search engines. Wayback Machine (Maltego Technologies) enables the exploration of more than 439 billion archived websites. Maltego also has some limitations. It takes moderate care of user security, requiring not only registration and authentication, but also processing requests on its own servers. Personal data is shared with additional modules and plugins. Maltego is an expensive tool if the user wants to use all functionalities (subscription). Using the free version of the tool involves limits (maximum of 12 results). In addition, Maltego focuses mainly on the IT aspects, neglecting the cultural and social context.

OSRFramework (Open Sources Research Framework)[60] – is a powerful toolkit for OSINT tasks, created in 2015 by Félix Brezo and Yaiza Rubio, cyber security researchers from Spain. This collection of libraries and scripts, written mainly in Python, automates the process of collecting and analysing publicly available information. The OSRFramework's modular architecture, which allows for extensions and customisation to meet specific needs, features good integration with other OSINT tools. Key functions include: checking the availability and existence of a username on different platforms, collecting information about email addresses, searching for information about a person or subject, about domains, about phone numbers, as well as creating a comprehensive profile of a person and generating reports from the collected data. Disadvantages of OSRFramework include a steep learning curve, dependence on external APIs, potential privacy issues and the possibility of generating false positives in search results.

SpiderFoot[61] – the tool was developed in 2012, released under the GNU General Public License v2.0, and was created by Steve Micallef, a security specialist and software developer. It was developed for cyber security reconnaissance and OSINT. It focuses on extracting information about domains, IP addresses, networks, blockchain exploration, SSL/TLS certificates, and to a limited

[60] OSRFramework, https://github.com/i3visio/osrframework [accessed: 28 VI 2024].

[61] Spiderfoot, https://github.com/smicallef/spiderfoot – installed version [accessed: 28 VI 2024], https://login.hx.spiderfoot.net/ – online version (HX) [accessed: 28 VI 2024].

extent performs searches of social media, discussion forums and the Darknet (for leaks). It consists of more than 200 different tools. It operates online (in a limited free version and a full paid version on a subscription basis) or as locally installed software (older version). The online version is called SpiderFoot HX. The undoubted disadvantage is that the software is primarily aimed at cyber security professionals, not white intelligence. It is a paid tool and the focus of the OSINT operators is visible to the software provider (HX version).

Recon-ng[62] – was created in 2013, its creator is Tim Tomes, also known as LaNMaSteR53. The tool is written in Python, and works as a framework where users utilize various modules to collect and analyse data:

- Recon - used to collect information from external sources, such as search engines, WHOIS databases, social media, etc.;
- Reporting - modules designed to create reports;
- Exploitation - exploration of potential vulnerabilities - less useful for white intelligence purposes;
- Import/Export - modules for importing and exporting data.

It is an open source tool, released under a BSD 3-Clause New/Revised license. Most Recon-ng modules are free, but some may use external services or APIs requiring paid access. You must be aware of the possibility of using download limits, and it only works on GNU/Linux systems. The tool runs in a terminal and requires a specific command language.

TheHarvester[63] – allows the collection of information on domains, email addresses, usernames, host names and IP addresses. It was created by Christian Martorell in 2007 and is maintained by the Edge Security team. It is an open source tool (GPLv3 license) available for use on your own computer. It allows searches using: search engines (Google, Bing, Baidu), social networks (LinkedIn, Twitter), WHOIS, SSL/TLS certificate registries, DNS. It runs in a terminal and requires learning a specific command language. It is intended for those coping with Bash/Python/GNU/Linux systems.

FOCA (Fingerprinting Organizations with Collected Archives)[64] – a web-based tool designed for open source intelligence. First appearing in 2009, it was developed by the Spanish cyber security company ElevenPaths. It was released under a GNU (FOCAFree) license. It is mainly used to find metadata and information hidden in scanned documents. Documents can be scanned online or locally. It identifies sensitive information such as network structure, user data

[62] Recon-ng, https://github.com/lanmaster53/recon-ng [accessed: 28 VI 2024].

[63] TheHarvester, https://github.com/laramies/theHarvester [accessed: 28 VI 2024].

[64] FOCA, https://github.com/ElevenPaths/FOCA [accessed: 28 VI 2024].

(names, webonyms), technical data and software versions. There is no search engine of its own, the document retrieval process is done via Google, Bing and DuckDuckGo. The programme essentially scans Microsoft Office/Open Office, pdf, AdobeInDesign and svg documents. It can also work as a plugin, as implemented in the CAT browser. Disadvantages of FOCA include: the need to install an SQL server, partial payment (FOCAPro), a high entry barrier, a steep learning curve, no updates for two years and restriction to metadata analysis only.

Links to software and other OSINT tools

This is a diverse and numerous group. In fact, each OSINT operator has its own collection of links. The collections characterised by the greatest number of accumulated links or popularity among researchers are indicated.

Bellingcat's Online Investigation Toolkit[65] – a collection of tools for researching and verifying information on the internet developed by Bellingcat. This set takes the form of an MSExcel spreadsheet, containing 14 tabs with logically grouped links, annotated with names and comments on the purpose and rules for using each tool. The main focus is on social media and registers. Despite careful organisation, the form of the Excel spreadsheet is not conducive to ergonomic working. Limitations include: no updates from June 2023, limited imageboard exploration tools, no discussion forum addresses and no Darknet exploration tools. The spreadsheet only acts as an intermediary – directing to external tools, without embedding them. There is a noticeable lack of information registers on Central and Eastern Europe and the Russian Federation. However, it is a collection that every OSINT analyst should have in their resource.

OSINT Framework[66] **–** is a website containing an interactive map of classified links to OSINT tools and websites. Built in a graph form, it provides easy access to the tool needed. The creator of the OSINT Framework is Justin Nordine, a specialist in information security. The project was initiated in 2015 as a tool for the security, intelligence and information analysis community and is well-known and often used in OSINT community. There is a Polish equivalent called Otwarte Źródła (Open Sources) based on Nordine's code. It supplements the data with those for Poland[67]. The provider of the website built it in JavaScript, it is possible that it collects the digital fingerprint of the OSINT operator. Much of the information in Otwarte Źródła (Open Sources) is out of date.

[65] Bellingcat's Online Investigation Toolkit, https://heystacks.com/doc/612/bellingcats-online-investigation-toolkit-bitlybcat (spreadsheet) [accessed: 28 VI 2024].

[66] OSINT Framework, https://osintframework.com [accessed: 28 VI 2024].

[67] Otwarte Źródła (Eng. Open Sources), https://osintframework.pl [accessed: 28 VI 2024].

Malfrats OSINT Map[68] – created by Malfrats Industries, is a continuation of the OSINT Framework. It was created as a result of dissatisfied investigators who noticed a lack of updates in the OSINT Framework. Malfrats OSINT Map contains 18 categories, including a tab for OSINT investigations in the area of forces and military operations. Additionally, it offers categories such as tools for web analytics, dark web monitoring and more. However, the platform has security related drawbacks (JavaScript).

Meta OSINT Chart[69] – is the result of five years of work by TropChaud, a professional white intelligence researcher and analyst. The project is a free and open aggregation of tools and resources, designed to support beginners in OSINT investigations. What sets MetaOSINT apart is its innovative form of data presentation in the form of a bubble chart, where the size of the bubbles reflects the popularity of a particular tool among OSINT operators, as measured by the number of indications from respondents. The MetaOSINT dataset is based on a survey conducted by the project's author of almost 30 lists of OSINT tools and resources, containing nearly 5,000 source links.

Awesome OSINT For Everything[70] – a categorised, comprehensive collection of links that covers virtually all aspects of OSINT. It includes tools for web analytics, information retrieval, social media analysis and other tools often not more widely known. It is distinguished by its inclusion of AI (artificial intelligence) tools and selected tools for generating temporary false identities on the internet.

The Ultimate OSINT Collection[71] – is a comprehensive collection of OSINT tools with a variety of categories such as web analysis, people search, social media analysis and more. Friendly for beginner researchers.

There are many collections of this type, the scope and content of which are largely similar[72].

[68] Malfrats OSINT Map, https://map.malfrats.industries [accessed: 28 VI 2024].

[69] MetaOSINT Chart, https://metaosint.github.io/learn-more [accessed: 28 VI 2024].

[70] Awesome OSINT For Everything, https://github.com/Astrosp/Awesome-OSINT-For-Everything [accessed: 28 VI 2024].

[71] The Ultimate OSINT Collection, https://start.me/p/DPYPMz/the-ultimate-osint-collection [accessed: 28 VI 2024].

[72] The rich collections are: OSINT4All (https://start.me/p/L1rEYQ/osint4all), OSINT Tools Lorando Bodo (https://start.me/p/7kxyy2/osint-tools-curated-by-lorand-bodo), Nixintel's OSINT Resource List (https://start.me/p/rx6Qj8/nixintel-s-osint-resource-list) and Verification Toolset created by Julia Bayer (https://start.me/p/ZGAzN7/verification-toolset). Freedomlab tool libraries (https://www.freedomlab.io/tools-for-hrds) and Haystack collections (https://heystacks.com – there are more than just OSINT tools and databases available there) are absolutely to be recommended.

Summary and conclusions

The article categorises and evaluates selected OSINT tools used to search for information, with particular emphasis on open source solutions. The most important thing seems to be to indicate the limitations of popular search tools in the context of the scope and area of their application. It is observed that in OSINT research, insufficient attention is paid to security aspects. It is often assumed that it is sufficient to use a virtual machine with a secure GNU/Linux distribution, such as Whonix or Linux Tails[73], as follows, relative adherence to recommendations for the use of virtual private networks (VPN). This approach seems both misguided and impractical. Advanced security-oriented GNU/Linux systems can be ineffective due to connection routing through the Tor network, whose exit nodes are identified and often blocked by numerous services, including OSINT tools. The distinctive digital fingerprint of the computer with a secure GNU/Linux distribution installed may hint to the searcher that it is of interest to the OSINT operator. It is worth recalling that already 15 years ago, cyber criminals used surveillance techniques to monitor the activities of officers investigating against them. An example of this practice is the Bayrob criminal group, which was involved in selling non-existent vehicles on the e-Bay platform. Members of the group used sophisticated tracking and control methods to gain an advantage over the law enforcement agencies involved in their investigations[74]. VPN server addresses are also traceable and some services may block them. OSINT research rarely recommends and practices the use of private proxy servers, which provide the greatest opportunities for hiding among web users. At the same time, there is a lack of in-depth analysis of the issues surrounding the leaving of digital fingerprints and the tools that allow this aspect of online identity to be freely shaped[75]. Also missing from the literature is an in-depth reflection on the relationship between security and search capabilities, where

[73] In the context of open source intelligence (OSINT), systems such as Linux Tails (https://tails.net) and Linux Whonix (https://www.whonix.org), despite their high level of security, can generate some functional limitations. It is worth considering the use of lesser-known distributions, such as Kodachi Linux (https://www.digi77.com/linux-kodachi/), Qubes OS (https://www.qubes-os.org) or Subgraph OS under development (https://subgraph.com/sgos/download/index.en.html), which offer a balance between security and freedom of intelligence activities.

[74] United States Court of Appeals, *United States of America v. Bogdan Nicolescu; Radu Miclaus,* https://www.opn.ca6.uscourts.gov/opinions.pdf/21a0231p-06.pdf [accessed: 28 VI 2024].

[75] The OSINT analyst can see the usefulness of tools such as Dolphin {Anty} web browser (https://dolphin-anty.com/en/), which allows the free shaping of digital fingerprint, or Ghacks set of advanced configuration settings for the Firefox browser, designed to improve privacy, security and performance. There are many other solutions that vary in scope and content, and an overview of these could be the subject of a separate text.

increasing security levels often limits search potential. It is common to uncritically adopt closed and commercial OSINT tools, ignoring the fact that the analyst himself becomes the object of the software provider's observation and that partial results of the investigation may be known to the provider. Therefore, the author of this article strongly recommends the use of tools autonomously owned by the OSINT operator. A significant problem is the imbalance between information acquisition and analysis. Issues related to intelligence analytics and issues of cognitive biases, heuristics and information noise often remain on the periphery of OSINT professional's interests. Paradigms of analysis and awareness of the possibility of cognitive biases are key issues, as important as the tools. Every open source intelligence analyst should have these issues mastered. Another challenge is the conscious selection of tools to support the analysis and visualisation of OSINT results. It is observed that researchers, following the trend, are limiting themselves to open source software, such as the open source CherryTree[76] or the commercial Hunchly[77]. Alternatively, they turn to online solutions such as Visual Investigative Scenarios (VIS)[78]. Security is also overlooked in the choice of tools, and commercial software and online services may be subject to surveillance. There is a noticeable lack of comprehensive review of the available software in terms of its impact on the security and confidentiality of OSINT analyses carried out[79]. Another problem is the marginal use of quantitative methods. Qualitative analyses or those based on colloquial inference dominate. It is also important to pay attention to new technologies that may prove to be breakthroughs in the field of OSINT, especially large language models, referred as to a generative AI. These tools can be used both for information retrieval and for information processing as well as analysis. The considerable dispersion of the OSINT software development community can be assessed ambivalently. The multitude of tools with similar functions and short existence cycle, ending in commercialisation or project abandonment, makes it difficult to choose the optimal solution. Commercial tools, despite their convenience of use, may limit the cognitive perspective of the OSINT analyst and raise privacy concerns. An open, modular tool with a low entry threshold and a gentle learning

[76] CherryTree, https://www.giuspen.com/cherrytree [accessed: 28 VI 2024].

[77] Hunchly, https://www.hunch.ly [accessed: 28 VI 2024].

[78] VIS, https://vis.occrp.org [accessed: 28 VI 2024].

[79] Trilium (https://github.com/zadam/trilium) is worth noting – it appears to meet the demands made in the text. This open source software, running on the user's computer, provides control over data confidentiality and offers a set of functions to support the collection and visualisation of OSINT investigation.

curve, allowing a wide range of users to actively participate in the development of the software and customise it, seems to be the optimal solution.

Based on the analyses and considerations carried out, two potential development scenarios towards OSINT 3.0 can be anticipated – a positive and a negative one.

The positive scenario assumes an increase in the integration of OSINT tools with operating systems and browsers, which would increase user convenience and operational security. Examples such as Trace Labs or CAT suggest that the future of OSINT could be based on deeper integration, allowing tools to be pre-installed and configured in secure environments. The development of advanced data analysis and visualisation mechanisms, such as interactive relationships maps, advanced social network graphs and dynamic predictive models, is also anticipated. The development of tools designed to explore specific web spaces, including the dark web, metaverses, decentralised networks and blockchains, can also be expected. An important element may be the automation of processes using artificial intelligence, which will increase the efficiency of OSINT operations and enable better management of large data sets.

The negative scenario predicts a trend towards fragmentation of OSINT tools. Instead of consolidation of functions, increasing specialisation and fragmentation of tools is possible, increasing the complexity of OSINT operator's work. Tools may become more complex and less intuitive. This will reduce their accessibility, especially for new users, and lead to a decrease in efficiency. An increasing reliance on commercial tools and databases is also possible. The disappearance of open and free solutions may drive users towards paid tools, increasing software provider's control over access to data while narrowing the operational options for smaller organisations and independent researchers.

In summary, the future of OSINT tool development remains unresolved. On the one hand, positive changes related to integration, automation and the development of advanced analytical functions are possible. On the other hand, there is a risk of fragmentation, complication of use and increasing commercialisation, which may negatively affect the availability and functionality of OSINT tools.

Bibliography

Abramczuk K., Kąkol M., Wierzbicki A., *How to Support the Lay Users Evaluations of Medical Information on the Web?*, in: *Human Interface and the Management of Information: Information, Design and Interaction*, S. Yamamoto (ed.), Cham 2016, pp. 3–13. https://doi.org/10.1007/978-3-319-40349-6_1.

Bazzell M., *Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information,* Charleston 2018.

Bazzell M., *OSINT Techniques: Resources For Uncovering Online Information*, [n.p.] 2023.

Block L., *The long history of OSINT*, "Journal of Intelligence History" 2023, vol. 23, no. 2, pp. 95–109. https://doi.org/10.1080/16161262.2023.2224091.

Dorn A.W., *United Nations Peacekeeping Intelligence*, in: *The Oxford Handbook of National Security Intelligence*, L.K. Johnson (ed.), Oxford 2010, pp. 275–295.

Forge J., *A Note on the Definition of "Dual Use*", "Science and Engineering Ethics" 2010, vol. 16, no. 1, pp. 111–118.

Hargittai E., Hinnant A., *Digital Inequality: Differences in Young Adults' Use of the Internet*, "Communication Research", 2008, vol. 35, no. 5, pp. 602–621. https://doi.org/10.1177/0093650208321782.

Hulnick A.S., *Fixing the Spy Machine. Preparing American Intelligence for the Twenty-First Century*, Westport 1999.

Lowenthal M.M., *Intelligence. From Secrets to Policy*, Washington 2007.

Maddrell P., *Spying on Science: Western Intelligence in Divided Germany 1945-1961*, Oxford 2006.

Mercado S.C., *Sailing the Sea of OSINT in the Information Age*, "Studies in Intelligence" 2004, vol. 48, no. 3, pp. 45–55.

Mider D., *Mappa Mundi ukrytego Internetu. Próba kategoryzacji kanałów komunikacji i treści* (Eng. Mappa mundi of the hidden internet. Categorising internet communication channels), "PTINT Praktyka i Teoria Informacji Naukowej i Technicznej" 2015, vol. 23, no. 1, pp. 3–16.

Mider D., *Sztuka wyszukiwania w Internecie – autorski przegląd wybranych technik i narzędzi* (Eng. The art of searching on the internet. Review of selected techniques and tools), "Studia Politologiczne" 2019, vol. 54, pp. 191–229.

Mider D., Garlicki J., Mincewicz W., *The Internet Data Collection with the Google Hacking Tool – White, Grey or Black Open-Source Intelligence*, "Internal Security Review" 2019, no. 20, pp. 280–300.

Nasheri H., *Economic Espionage and Industrial Spying*, Cambridge 2004.

National Research Council, *Computers at Risk: Safe Computing in the Information Age*, Washington 1991.

Olcott A., *Open Source Intelligence in a Networked World (Continuum Intelligence Studies)*, New York 2012.

Open Source Intelligence Market Size, Share, Competitive Landscape and Trend Analysis Report by Source, Technique and End User: Global Opportunity Analysis and Industry Forecast, 2020-2027, Allied Market Research, May 2020.

Rosenzweig P., McNulty T.J., Shearer E., *Whistleblowers, Leaks, and the Media: The First Amendment and National Security*, Chicago 2013.

Schaurer F., Störger J., *Guide to the Study of Intelligence. The Evolution of Open Source Intelligence (OSINT)*, "The Intelligencer: Journal of U.S. Intelligence Studies" 2013, no. 3, pp. 53–56.

Steele R.D., *Open source intelligence*, in: *Handbook of Intelligence Studies*, New York 2007, pp. 129–147.

Steele R.D., *The Open-Source Everything Manifesto: Transparency, Truth, and Trust*, Berkeley 2012.

Turaliński K., *Wywiad gospodarczy i polityczny. Podręcznik dla specjalistów ds. bezpieczeństwa, detektywów i doradców gospodarczych* (Eng. Economic and political intelligence. A handbook for security professionals, investigators and economic advisers), Warszawa 2015.

Tylutki K., *The information of a mass destruction range – OSINT in intelligence activities,* "Internal Security Review" 2018, no. 19, pp. 384–404.

Wyniki pracy wywiadu naukowo-technicznego MSW PRL 1971–1989 (Eng. The results of the operations of the scientific-technical intelligence of the Polish People's Republic 1871-1989), M. Sikora (comp.), Katowice–Warszawa 2019.

Internet sources

A Consumer's Guide to Intelligence, Office of Public Affairs CIA, 1999, https://archive.org/details/consumersguide_tenet/mode/2up [accessed: 28 VI 2024].

AFP, *How Bellingcat became Russia's 'biggest nightmare'*, France24, 7 IX 2022, https://www.france24.com/en/live-news/20220907-how-bellingcat-became-russia-s-biggest-nightmare [accessed: 28 VI 2024].

Awesome OSINT For Everything, https://github.com/Astrosp/Awesome-OSINT-For-Everything [accessed: 28 VI 2024].

Bellingcat's Online Investigation Toolkit, https://heystacks.com/doc/612/bellingcats-online--investigation-toolkit-bitlybcat (spreadsheet) [accessed: 28 VI 2024].

BlackArch, https://blackarch.org/index.html [accessed: 28 VI 2024].

BrowserAudit, https://browseraudit.com [accessed: 28 VI 2024].

CherryTree, https://www.giuspen.com/cherrytree [accessed: 28 VI 2024].

Chertoff M., Simon T., *The Impact of the Dark Web on Internet Governance and Cyber Security*, https://www.cigionline.org/static/documents/gcig_paper_no6.pdf [accessed: 28 VI 2024].

Ciancaglini V. et al., *Deep Web and Cybercrime: It's Not All About TOR*, Trend Micro, 12 XI 2014, https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/deep-web-and-cybercrime-its-not-all-about-tor [accessed: 28 VI 2024].

Colquhoun C., *A Brief History of Open Source Intelligence,* Bellingcat, 14 VI 2016, https://www.bellingcat.com/resources/articles/2016/07/14/a-brief-history-of-open-source-intelligence/ [accessed: 28 VI 2024].

Electronic evidence of war crimes. The role of journalists, media and social media, webinar organised by Group of Friends on the Safety of Journalists and Media Freedom in Strasbourg and the Council of Europe, 25 XI 2022, https://www.coe.int/en/web/kyiv/-/electronic-evidence-of-war-crimes-and-the-role-of-journalists-media-and-social-media [accessed: 28 VI 2024].

FOCA, https://github.com/ElevenPaths/FOCA [accessed: 28 VI 2024].

Higgins E., *How Open Source Evidence was Upheld in a Human Rights Court*, BellingCat, 28 III 2023, https://www.bellingcat.com/resources/2023/03/28/how-open-source-evidence--was-upheld-in-a-human-rights-court/ [accessed: 28 VI 2024].

Hunchly, https://www.hunch.ly [accessed: 28 VI 2024].

IntelTechniques, https://inteltechniques.com [accessed: 28 VI 2024].

Kali Linux, https://www.kali.org [accessed: 28 VI 2024].

Mackinnon A., *Bellingcat Can Say What U.S. Intelligence Can't,* Foreign Policy, 17 XII 2020, https://foreignpolicy.com/2020/12/17/bellingcat-can-say-what-u-s-intelligence-cant/ [accessed: 28 VI 2024].

Malfrats OSINT Map, https://map.malfrats.industries [accessed: 28 VI 2024].

Maltego, https://www.maltego.com [accessed: 28 VI 2024].

Market share of leading desktop search engines worldwide from January 2015 to January 2024, Statista, 2024, https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed: 28 VI 2024].

Matthews O., *Fact Cats. The inside story of how it got the Skripal scoop,* The Spectator, 20 X 2018, https://www.spectator.co.uk/article/fact-cats/ [accessed: 28 VI 2024].

MetaOSINT Chart, https://metaosint.github.io/learn-more [accessed: 28 VI 2024].

NATO Open Source Intelligence Handbook v 1.2, https://archive.org/details/NATOOSINTHandbookV1.2/page/n1/mode/2up [accessed: 28 VI 2024].

Open Source Intelligence Market Size, Share, Growth, and Industry Analysis, By Type (Video Analytics, Text Analytics, Visualization Tool, Cyber Security, Web Analysis, Social Media Analysis, and Others), By Application (Private Sector, Public Sector and Other), Regional Insights, and Forecast to 2032, Business Research Insights, March 2024, https://www.businessresearchinsights.com/market-reports/open-source-intelligence-market-109546 [accessed: 28 VI 2024].

OSINT Framework, https://osintframework.com [accessed: 28 VI 2024].

OSRFramework, https://github.com/i3visio/osrframework [accessed: 28 VI 2024].

Otwarte Źródła (Eng. Open sources), https://osintframework.pl [accessed: 28 VI 2024].

Paley N., *Copying is an act of love. Please copy and share*, https://copyheart.org [accessed: 28 VI 2024].

ParrotOS Security, https://www.parrotsec.org [accessed: 28 VI 2024].

PrivacyTests, https://privacytests.org [accessed: 28 VI 2024].

Recon-ng, https://github.com/lanmaster53/recon-ng [accessed: 28 VI 2024].

Reinsel D., Grantz J., Rydning J., *The Digitization of the World. From Edge to Core*, https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf [accessed: 28 VI 2024].

Spiderfoot, https://github.com/smicallef/spiderfoot [accessed: 28 VI 2024].

Spiderfoot, https://login.hx.spiderfoot.net [accessed: 28 VI 2024].

The Ultimate OSINT Collection, https://start.me/p/DPYPMz/the-ultimate-osint-collection [accessed: 28 VI 2024].

TheHarvester, https://github.com/laramies/theHarvester [accessed: 28 VI 2024].

Trace Labs, https://www.tracelabs.org/initiatives/osint-vm [accessed: 28 VI 2024].

Tsurugi Linux, https://tsurugi-linux.org/index.php [accessed: 28 VI 2024].

VIS, https://vis.occrp.org [accessed: 28 VI 2024].

Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025, Statista, June 2021, https://www.statista.com/statistics/871513/worldwide-data-created/ [accessed: 28 VI 2024].

White E., *Closing cases with open-source: Facilitating the use of user-generated open-source evidence in international criminal investigations through the creation of a standing investigative mechanism*, Cambridge University Press, 7 IX 2023, https://www.cambridge.org/core/journals/leiden-journal-of-international-law/article/closing-cases-with-opensource-facilitating-the-use-of-usergenerated-opensource-evidence-in-international-criminal-investigations-through-the-creation-of-a-standing-investigative-mechanism/981CEFF9D5AF80B6FD0A75BE6A1A384C [accessed: 28 VI 2024].

Williams H.J., Blum I., *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*, RAND, 17 V 2018, https://www.rand.org/pubs/research_reports/RR1964.html [accessed: 28 VI 2024].

Case law

Case of Ukraine and the Netherlands v. Russia, 8019/16, 43800/14, 28525/20, Archive of the European Court of Human Rights, 30 XI 2022, https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-222889%22]} [accessed: 28 VI 2024].

United States Court of Appeals, *United States of America v. Bogdan Nicolescu; Radu Miclaus,* https://www.opn.ca6.uscourts.gov/opinions.pdf/21a0231p-06.pdf [accessed: 28 VI 2024].

Daniel Mider, Associate Professor

Assistant Professor at the Faculty of Political Science and International Studies, University of Warsaw. He specialises in the issues of open source intelligence on the internet, crypto-assets, cybercrime, the sociology of internet and the sociology of political violence.

Contact: d.mider@uw.edu.pl