SECURITY SERVICE OF UKRAINE

Institute of Security Service of Ukraine
of Yaroslav Mudryi National Law University

HOW TO USE OSINT TOOLS AND METHODS FOR ACQUIRING NEW INFORMATION

Practical guide
5th edition, revised and extended

Харків
2025

УДК 343.3, 004.9 (06)

Recommended for publication by the Academic Council of the Institute of Security Service of Ukraine of Yaroslav Mudryi National Law University (protocol № 30, dated 23 April 2024)

Reviewers:
VOLODYMYR KARASTELOV,
officer of the Department of Counterintelligence Protection of State Interests in the Field of Information Security of the SSU, PhD in Law;
OLEKSANDR ZHUPINA,
officer of the General Investigation Department of the SSU, PhD in Law.

Authors:
DMYTRO ZORENKO,
associate professor of the Department of the Institute;
LIUDMYLA KULCHYTSKA, officer of the SSU;
ROMAN LEKH, deputy director (for educational and scientific work) of the Educational and Research Institute of State Security of the SSU National Academy, PhD in Law;
OLEKSANDR CHERVIAKOV, head of the Institute, PhD in Law.

Zorenko D. S., Kulchytska L. O., Leh R. V., Cherviakov O. I.
How to Use OSINT Tools and Methods for Acquiring New Information : Practical Guide. 5th edition, revised and extended. Kharkiv: Institute of SSU, 2025. 80 p.

Зоренко Д. С., Кульчицька Л. О., Лех Р. В., Червяков О. І.
Використання інструментів та методів OSINT для отримання пошукової інформації : практичний порадник. 5-те вид., переробл. та доповн. / Д. С. Зоренко, Л. О. Кульчицька, Р. В. Лех, О. І. Червяков. – Харків. Видавець: О. А. Мірошниченко, 2025. – 80 с.

ISBN 978-617-8130-72-5.

The practical guide comprehensively provide instructions on the theoretical and applied aspects of using OSINT (open source intelligence) tools and methods to search the Internet for information about persons, facts or events in order to meet the needs of the SSU bodies and units, as well as to develop the relevant digital competencies of the Service's employees.

This material is intended to develop basic knowledge and skills of the user and does not describe the ways of unauthorized interference with the operation of electronic computers, automated systems, computer or telecommunication networks. When considering search tools, the emphasis is placed on publicly available and free solutions, and the list provided is not exhaustive. The functionality of the web resources or programs specified in this guide may change over time.

The authors are not responsible for any damage or losses that may be incurred by the user (third parties) as a result of misunderstanding and/or application of the information provided in the publication. Web researchers must independently ensure the protection of their own personal data and other confidential information during search activities on the Internet. Any information obtained by the user applying the material presented herein is used at their own risk. Images taken from open sources are utilized in the design of this publication.

The guide is interded for employees of the SSU operational and investigative units, higher education students, teachers and researchers of departmental educational institutions, as well as anyone interested in the current issues of searching for information in open sources.

УДК 343.3, 004.9 (06)

ISBN 978-617-8130-72-5

CONTENTS

Symbols:
Bellingcat – hyperlink to the web resource;
bot – Telegram bot;
GitHub – open source project repository;
free – the scope of free functions on an online service or program;
RU (BY) – the resource is directly or indirectly related to the rf (rb).

1. The Concept and Purpose of OSINT

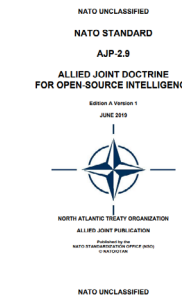
OSINT (*Open Source INTelligence*) is a component of special activities that involve obtaining information from open sources (publicly or commercially available), processing, analyzing and disseminating it. It is used to make decisions in the field of national defense and security, investigations etc.

OSINT as an individual practice started in the Unites States in the 1940s the same time as the Foreign Broadcast Monitoring Service was founded. Staff members of this service were recording and analysing foreign radio broadcasts, after that acquired data was given to the military and intelligence organs as reports. According to the CIA and Pentagon experts, during the «cold war» the United States received 70-90% of its intelligence information from open sources and only 10-30% from agents. Before the Internet, OSINT was based on analysis of print media, tele- and radio casts, photo- and video materials, various documents and report, research projects, inventions etc.

Nowadays open source intelligence is successfully used not only by the security and defence agencies of the world`s leading countries, but also by commercial companies, non-governmental organizations, analytical centres, journalists, different kinds of investigators, private individuals etc.

The most systematic descriptions of the content and procedural content of OSINT are presented by *NATO* experts in the publications «NATO Open Source Intelligence Handbook» (2001), «NATO Open Source Intelligence Reader» (2002), «NATO Intelligence Exploitation of the Internet» (2002), and also in the standards «Allied Joint Doctrine for Open-Source Intelligence» (AJP-2.9, 2019) and «Open-Source Intelligence (OSINT) Tactics, Techniques and Procedures» (AIntP-22, 2022).

AJP-2.9 and AIntP-22 define a set of commonly recommended tactics, techniques and procedures for implementing a standardized OSINT process at the operational level in support of NATO operations. They describe the role of the intelligence cycle as a key methodology for collecting and analysing search information in the execution of Joint Intelligence, Surveillance, and Reconnaissance activities. While these standards are intended primarily as guidance for NATO Joint Command and Headquarters, they can be used by non-allied partners and serve as a practical reference for the civilian personnel. OSINT delivers a comprehensive contribution to the intelligence cycle.


NATO UNCLASSIFIED

NATO STANDARD

AJP-2.9

ALLIED JOINT DOCTRINE FOR OPEN-SOURCE INTELLIGENCE

Edition A Version 1

JUNE 2019

NORTH ATLANTIC TREATY ORGANIZATION

ALLIED JOINT PUBLICATION

NATO UNCLASSIFIED

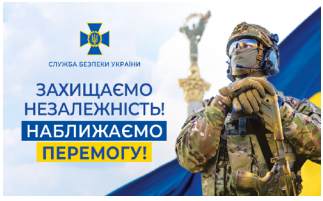
On the other hand, the public sector and online media are actively integrating OSINT technologies into the practice of conducting public investigations of high-profile social and political events, corruption of high-ranking officials, the situation in armed conflict zones, human rights violations, and crimes against peace and security. One of the most successful examples is the work of the international team of journalists of the Bellingcat project (in particular, cases of the MH17 flight downing, shelling of Ukrainian territory by the Russian army in 2014-2015, identification of Wagner PMC militants involved in war crimes).

Given such a powerful response, the Ukrainian information space has rapidly gained momentum in the OSINT movement – InformNapalm, Molfar, OsintFlow, OSINT Bees, Truth Hounds – and this is a far from complete list of OSINT communities that have been involved in helping the defence forces, collecting evidence of war crimes, and countering enemy propaganda since the beginning of russia's armed aggression against Ukraine. The activists also believe that one of their tasks is to systematize the modern experience of OSINT research for further systematic implementation in the activities of interested government agencies.

Undoubtedly collecting and analysing open data is an important component of countering current challenges and threats to Ukraine's national security: thanks to OSINT tools and methods, specialists of SSU identify the location of enemy military formations and vehicles, identify personal data of war criminals and collect the appropriate evidence, expose betrayers and collaborators, initiate the inclusion of individuals or organizations in sanctions lists, use psychological and informational influence on the enemy, expose fakes of enemy propaganda etc.

Today OSINT is not just a desk job, but a highly applied and effective discipline that allows us to obtain information about enemy military objects, plans, and documents without risking the lives of employees. At the same time, it requires enormous resources and time to analyse the data and determine its impact on the security of our country. In this process, the SSU effectively cooperates with national and foreign partners who provide advanced software systems and informational support.

Since search activities using open sources significantly increase the institutional capabilities of the Ukrainian special services, OSINT has become one of the required elements of professional training of SSU employees due to a successful combination of the best available international practices and unique practical experience gained during the fight against russian armed aggression.



2. Stages of Research (Intelligence Cycle)

NATO standards focus on the necessity to follow a certain algorithm during the search work – the so-called Intelligence Cycle, which consists of the following **TCPED stages**: Task, Collect, Process, Exploit, and Disseminate. This can be accomplished in a deliberate, ad hoc or dynamic time frame in support of operations planning and execution.

1) Task – defines OSINT research tasks by the requesting party (for example, identification of an individual or legal entity, finding out certain aspects of its activities, close environment or related contacts, compromising materials, location of an object or its movement route, verification of available data etc.), determining the necessary resources and methods for its implementation, and providing this information to the direct executor.

The Executor should analyse such requested information for its compliance with the criteria of specificity, measurability, relativity, verifiability and urgency; understand its scope and nature, deadlines and forms of reporting; determine the sufficiency of resources for its performance; the level of potential threats to the safety of data collection and possible legal reservations. If necessary, the Executor may contact the ordering party to clarify the details of the task or to provide additional background information.

2) Collect consists of: 1) developing an optimal data collection methodology (preparing a plan, identifying the necessary methods, tools and resources; preparing a list of keywords for search requests, risk management etc.); 2) searching for information using known and processing previously unknown or alternative sources; 3) directly collecting relevant data. The process of evaluating the reliability of these sources and the accuracy of the information contained as well as their classification into *primary* (reports of individuals who had direct contact with the information) and *secondary* (citation or use of the primary source in any other way) is very important.

The final stage is focused collection, which may lead to extended collection, depending on the set timeframe.

The *5W+H scheme* can be considered as a basic model for research planning (Who, What, When, Where, Why and How) – then it becomes clearer where and in what form the necessary information can be stored. The search process is like filling in the segments of a mosaic - each new reliable piece of information is immediately used in the search work to expand its effectiveness. It is difficult to define a universal set of tools and an algorithm for search actions

– keywords, sources of potential information, the necessary tools, the availability of information and its reliability – depend on the source data and the end goal of the search.

Potential sources of information on the subject of interest may include: *a person* (social media, channels, blogs, forums etc.); *his or her environment* (family, friends, neighbours, employer, colleagues, employees, potential competitors); *the state, local authorities, companies, institutions or organizations* (registers, databases, court decisions, debts, official correspondence, online media etc.).

The data collection can be *passive* (without interaction with the target of interest, for example, using search engines, services or Telegram bots) and *active* (for example, using social engineering elements when working with email or social media accounts; scanning for vulnerabilities in the operating system and web applications; gaining access to open ports on devices etc.).

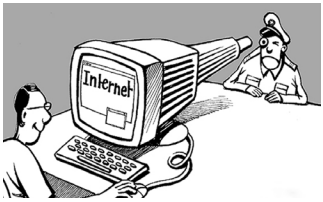
Since data on the Internet can be deleted by the owner at any time, it is a good idea to create an *offline archive of the collected information* (e.g., using screenshots; creating pdf files with the website url, date, time and title; saving the entire web page or individual files; archiving websites etc.). Any result collected during this stage can trigger an iteration of the collection step.

3) Process is the analysis of all collected information for relevance (compliance with the research goal) and its sorting according to certain criteria (object/content/source/time/reliability/connection etc.); translation and adaptation of foreign language materials; adding standardized metadata to certain files; consolidation of information into a summarized data set by comparing and grouping related elements.

4) Exploit – verification of pre-processed information, i.e. comparing it with other independent OSINT sources to confirm or deny the reliability of the data. If such verification is not possible, the information is marked as «unverified». The final OSINT result is formed by combining *qualitative* (in-depth analysis of the subject area of research with the possible addition of conclusions, assessments, expert comments, explanations, and recommendations) and *quantitative* (use of visualization – diagrams, mind maps, infographics, screenshots/photos, extracts, links to sources, applications) processing of information, taking into consideration the relevance of the request.

5) Disseminate is preparation and delivering the final document to the client with possible warnings regarding its sharing (access restrictions, content, views or opinions expressed, purposes of use, liability, etc.). The client should evaluate the report, identify the problematic aspects of the research and suggest the ways to solve them (ensure that questions, problems, and concepts applied during all steps are evaluated).

3. Anonymization of Internet Search

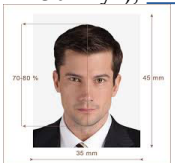


The Internet is a public space that is always evolving, and therefore there are no absolutely guaranteed ways to remain 100% anonymous. Any anonymity is temporary, and it can become harder to identify a particular user. In other words, the search must be carried out with the knowledge that your actions may potentially be recorded and analysed by third parties for the purpose of further identification of the researcher or the focus of his or her search work. A balanced approach would minimize unauthorized disclosure of sensitive data.

There is *social anonymity* (a person consciously abstains from disclosing personal information on web resources, for example, when registering on social media platforms) and *technical anonymity* (using special software solutions to mask / change the data of the device from which the search is carried out). Therefore, **it is necesary to separate your work activities from private while surfing the Internet** – personal accounts (browser operating system and other programs, email accounts, social networks or messengers) and equipment (laptop, tablet, mobile phone) should not be used for professional tasks, and vice versa. Experts state that anything that gets into the global network stays there forever. Everyone is responsible for ensuring the security of an investigation and those affected by it.

So, for each software product a researcher works with, as well as for working on web resources, messengers or social media that require the user to log in to access content or certain functionality, it is recommended to **use a virtual personality** (or several) that should imitate a real person to the greatest extent possible for the needs of a particular study (this can be a nickname, gender, age, region, biography, interests and hobbies, circle of friends, photos, etc.). Photographs, telephones, emails or data that are personal should never be used. Such a fake profile (sockpuppet) can consist of an any combination of the following elements:

• personal data (a fairly common name and surname, date of birth, country, place of residence, gender, etc.), that can be *made up* or *generated* – Busines, FakeDetails, FakeInfo, FakeNameGenerator, FakePersonGenerator, Meragor, miniRANDOM, NameFake, OnlineNameGenerator, RandomUserGeneration, Randus (RU, free – 10 request for 30 days), ThisResumeDoesNotExist, uk-osint.com;



• photo – BoredHumans, FaceApp, GeneratedPhotos (free – 3 days), Meragor (RU), RandomFaceGenerator (digital watermark is present), ThisPersonDoesNotExist (digital watermark is present), Unreal Person, WhichFaceIsReal, human face generator, portrait

generator, Kandinsky (RU, bot); AIFaceswap (creating deepfake photos and videos).

Currently, neural networks are best at close portrait photos, where the focus is on the face and the background is usually blurred. The main areas of mistakes in such photos are eyes, teeth, earlobes, hair, clothing or accessories, and writings. Also, pay attention to the fingers, they may have more than five and/or have unnatural bends.

Since such images are created by artificial intelligence following certain rules (for example, the eyes should always be on the same parallel, and the distance between the eyes and between the eyes and the mouth should be the same), it is recommended to check them both visually and with the help of special services. Therefore, before uploading the generated photo to a social network, it is considered the best practice to make slight adjustments to it using a photo editor (e.g., changing facial features, hairstyle, clothing, hair colour, adding glasses, etc.);

• email (without linking to a mobile phone number or other email) – registration of a *«permanent» address,* for example, on the resources of Addy.io (free – encryption of outgoing emails, mail forwarding function, 2 anonymous mailboxes and 1 real recipient mailbox, monthly mail traffic of 10 MB), Gmail (the requirement to link a phone number is constantly changing, currently it is not necessary when creating one mailbox via the web interface and without any unusual user activity; free – 15 GB storage), Mailfence (free – encryption, storage 1 GB), ProtonMail (positively perceived by social networks during registration; free – supports encryption of email content, unlike its subject and metadata, no logging, 1 GB storage, free VPN), Tuta (offers a code to restore access, free – end-to-end encryption, removal of IP address from emails, 1 GB storage).

Services such as NameCheckup та NaMint will help you to choose nicknames for your email and social media accounts (by first name, last name, and date of birth).

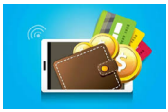
The created profiles should be periodically checked for leaks (for example, through DeHashed, Have I Been Pwned, Hudson Rock), and if necessary, the password should be changed (preferably at least 12 characters, including uppercase and lowercase Latin letters, numbers, and special characters, or use a password generator or similar), two-factor authentication should be enabled, or a new one should be registered.

Similarly, *a temporary (one-time) email* address has both its advantages (automatic generation of a random mailbox name, no link to any personal information) and disadvantages (most web resources do not accept it for account registration; each temporary email address is unique and issued only once, upon expiration of the term of operation, all data is deleted, so further confirmation of identity through it will be impossible; there are no guarantees that only one person – you – will have access to such mail). Disposable email addressing allows a different and unique email address for every sender or recipient combination. Ideally, owners share a disposable email address once with each contact or entity.

The most popular services of this kind include DropMail (lifespan – until the

web page is updated, built-in mail forwarding function, ability to restore the mailbox but not its contents), Gmailnator (mail generator @gmail.com, lifespan – 10 min), Guerrilla Mail (lifespan – 60 min, sends emails with attachments up to 150 MB, password generator with the possibility of storing and restoring them by master code), mail.tm (lifespan – until deleted, receives mail only, has a password), Tempr.email (lifespan – up to 30 days, receives and sends mail with attachments, access via password); Temp-mail and 10minutemail (lifespan – 10 min, can be extended up to 1 hour, you can reply to the received email or redirect it to another mailbox);

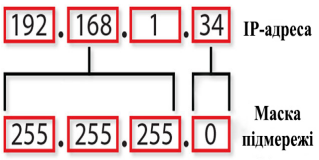
• a virtual phone number (for receiving SMS during a one-time registration in social networks/messengers or long-term lease) – FreeOnlinePhone, GetFreeSMSNumber, ReceiveSMSOnline, Sellaite.com (limited number of numbers, multiple use, possibility of access by other users, SMS responses are visible to all participants). However, it is better to use *paid resources* such as GrizzlySms (RU), OnlineSIM (RU), Proovl, Receive-SMS, Sms-Activate (RU), SmsHub (RU), Sms-reg (RU), TempNumber, Twilio or *marketplaces of ready-made social media accounts, email or messengers* – Accsmarket (RU), AccountsStore (RU), Buyaccs.com (RU), DarkStore (RU), Install-shop (RU), Olimp-shop.net (RU), etc.



• credit card number (and related personal data) – CardGeneration, CardGenerator, CardGuru, Fake Credit Card Number Generator, VCCGenerator.

Important: payment for social media accounts, VPN, Telegram bots, mobile operators (Bitrefill service) or any other advanced functionality of web services is made through a non-custodial cryptocurrency wallet.

In order to create an effective system of technical security for online search, it is necessary to understand the **potential threats to anonymity** (or vulnerabilities of the online environment) and know of the ways to minimize them. The basic ones include:



a) IP address (Internet Protocol Address) a unique number of a device in a computer network used to address data transmission. It is usually a combination of 4 numbers from 0 to 255, separated by periods (for example, 192.168.0.154 or 203.113.89.134). When you browse a website or connect to a network computer, you are accessing a specific IP address. A meaningful name for an online resource or device (for example, rada.gov.ua or user) is simply a way to display this IP address so that people can remember it more easily. In the IPv4 version of the protocol, the IP address is 4 bytes long, and in the IPv6 version, it is 16 bytes long. An IP address consists of two parts: a network number and a host number.

An IP address is linked to a meaningful name using the DNS (Domain Name System). When a user enters the name of a website, such as google.com, in the

address bar of a browser, the computer requests the IP address of that website from a special DNS server and, upon receiving the correct answer, opens the website itself. A DNS server is a specialized computer (or group of computers) that stores IP addresses of websites according to their meaningful names and processes user requests. There are a lot of DNS servers on the Internet, and each provider has them to deliver services to their subscribers.

The IP address of a device can be *statically* determined by the network administrator (provider) or *dynamically* assigned each time it connects via the DHCP (Dynamic Host Configuration Protocol, which allows a computer to automatically receive the parameters necessary for networking).

Also, all IP addresses are divided into *private* (so-called grey, used within a local/home network or LAN) and *public* (so-called white, intended for addressing on the World Wide Web or WAN). For example, if you have a Wi-Fi router in your apartment, each device (computer, smartphone, TV, etc.) is usually dynamically connected to its private network and receives a grey IP address for internal identification. You can find out the IP address of such a computer in the router settings or in the operating system (OS) using the console command *ipconfig /all* (OS Windows), *ifconfig en0* (MacOS) or *ifconfig -a* (Unix OS). To access the Internet, devices on a private network (with grey IP addresses) use a public IP address (white) provided by the provider using the NAT mechanism (network address translation, which replaces a private IP address with a public one). The router translates (replaces) the packet's return IP address with its external (visible from the Internet) IP address on the fly, and also changes the port number.

Similarly to a postal code, a user's public IP address allows you to determine the name of the provider and its location – 2ip.ua, Browserleaks, CentralOps, Deviceinfo, ipapi, ipleak.net, ip-score.com, Whoer (RU), Whois. At the same time, data on access to a particular web resource is constantly transmitted to the corresponding DNS server (for example, Dnsleak, Dnsleaktest, ViewDns.info). Therefore, masking your public IP address allows you to avoid tracking, bypass territorial restrictions and visit blocked websites. The most common ways to *hide* your IP address are to use a *VPN*, *proxy server*, *Tor browser*, or *public Wi-Fi*.

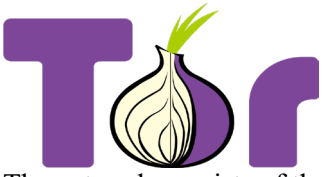
VPN (Virtual Private Network) is a network architecture for virtually extending a private network (i.e. any computer network which is not the public Internet) across one or multiple other networks which are either untrusted (as they are not controlled by the entity aiming to implement the VPN) or need to be isolated (thus making the lower network invisible or not directly usable). The security of information transmission via public networks is realized by means of encryption, which creates a channel of information exchange closed to outsiders.

Main advantages: *anonymity in the network* (Internet traffic is fully encrypted and inaccessible to third parties, for example, a provider; instead of a real IP address, the IP address of the VPN server is used; Similarly, DNS), *blocking bypass* (by changing the IP address, you change the geolocation, which allows you to bypass the blocking of access to Internet resources in a particular country or from a specific IP address), *protection against data interception and theft* (when visiting suspicious sites or connecting to free Wi-Fi networks, user traffic can be intercepted and sensitive data – metadata, passwords, cookies, sessions, etc. – stolen due to the lack of encryption and/or the implementation of a Man-in-the-Middle or MITM) hacker attack scenario).

There are *paid* and *free* VPNs, in the form of an *operating system application* (which is preferable, since all device traffic goes through the VPN server) or *a browser extension* (only connecting a specific browser).The fee-free nature of such services (e.g. BrowsecVPN, hide.me, ProtonVPN, TunnelBear, UrbanVPN) is achieved by setting restrictions on the amount of traffic or channel bandwidth, the number of available servers (countries), the inability to download torrents or stream video, and, most importantly, a non-transparent privacy policy.

It is better to choose a VPN that has strong encryption (e.g., AES-256), support for trusted security protocols (OpenVPN, L2TP, IKEv2, WireGuard), and protection against DNS leaks (in case your provider sends DNS queries regardless of whether the VPN is enabled), TrustedServer technology or its equivalent (deletes all your data every time you restart) and Kill Switch function or its equivalent (disconnects the Internet connection in case of loss of connection with the VPN server, prevents the disclosure of the real IP address). Therefore, paid versions of VPNs (e.g., CyberGhostVPN, ExpressVPN, Mullvad VPN, Private Internet Access, SurfShark VPN) have a much better level of protection, higher connection speeds, support for more locations or connected devices, multi-platform compatibility, anonymity of payment and accounts, no activity logs, etc. However, VPNs do not protect against trackers and do not affect cookies that have already been downloaded.

A proxy server (e.g., Anonymouse.ws, free-proxy.cz, Kproxy, ProxyScrape, Spys.one) is a mediating server for data exchange between the user and the target web resource. Although a proxy server replaces a person's real IP address with its own, unlike a VPN, it does not encrypt traffic and does not hide the researcher's actions on the global network. In the early days of the Internet, proxy servers were the most popular way to access the Internet from local networks.



Tor Browser (The Onion Router) is a secure version of the Firefox browser that accesses the Internet using its own network of anonymous proxy servers and virtual tunnels between them (Tor Network) to ensure that data is encrypted on all stages. The network consists of thousands of servers run by volunteers.

The onion analogy is based on the following mechanism implemented in Tor.

Each packet of information, entering this network, passes through three different proxy servers (nodes), which are selected randomly each time. Before being sent, the packet is sequentially encrypted with three keys: first for the third node, then for the second, and finally for the first. When the first node receives the packet, it decrypts the top layer of the cipher (analogous to peeling an onion) and finds out where to send the packet next. The second and third servers do the same. Within the Tor network, traffic is redirected from one router to another and finally reaches the exit point, from which a clean (unencrypted) data packet reaches the original recipient's (server's) address. In the opposite direction, traffic from the recipient is sent to the exit point of the Tor network. Moreover, the route is changed randomly every 10 minutes. As a result, no one has access to either the content or the address of the message. Its full decryption takes place only on the recipient's server. The disadvantage of this approach is the relatively slow operation of the browser. Keep in mind that one of the nodes in the Tor chain may well be vulnerable.

The greatest effect is achieved by *combining VPN and Tor*: in addition to hiding the IP address and encrypting traffic, to ensure anonymity, the browser does not store browsing history, cookies, or any information about the pages visited; all users have the same digital fingerprints.

When you connect to a public Wi-Fi network to access the Internet, a positive effect is achieved by using the «white» IP address of such a network (as opposed to the IP address of, for example, your home provider or your own smartphone) in combination with the simultaneous operation of an indefinite number of other subscribers. At the same time, it is considered risky because it does not protect against hacker attacks, viruses and other cyber threats. It is better to use it as a last resort, having previously launched a reliable VPN and antivirus.

b) cookies are fragments of data for further user identification that a website saves on your computer every time you visit a particular resource using your browser. After a person enters the address of a particular web page, the browser searches the device for the cookie of that site and, if it finds it, sends it to the resource's server. The site «recognizes» the user and automatically adjusts – registration forms will be filled in, language and regional preferences will be set. If the browser does not find cookies, the site considers you a new visitor and asks for permission to save such files.

It should be noted that cookies, cache and autocomplete are quite different technologies. A *cache* is a copy of large website data stored on a device (for

example, images, videos, and music). During the next visit to the site, the browser will not request this information again, but will take it from the cache, which will make the web page load faster. *Autofill* is a browser function. It remembers the data entered by the user when filling out forms on the site (name, phone number, email, etc.) and, if you need to fill out similar forms on another web page, offers these saved options.

Cookies are used for: *session management* (login, IP address and location of the user; date and time of visit to the site, OS and browser version), *personalization* (language, currency, font size or page scale), *behavioural tracking* (clicks and transitions, timing of stay on the site, viewed products or advertisements, selected sorting filters, preferences, likes, entered text – phone, email, bank card, comments, etc.).

They are generally divided into: *session* (located in the RAM and automatically deleted when the tab is closed) and *persistent* (stored on the device until a certain date or for a certain period); *own* (created directly on the site that the user opened) and *third-party* (the site contains clickable material from other resources such as banners or browser scripts, for example, Google Analytics, Facebook, Google AdSense, etc.); *necessary* (required for the normal functioning of the website); *zombie cookies* (flash cookies or super cookies, these are third-party persistent cookies that have a unique ability to recover after deletion; with their help, websites can block individual users or spread malicious software code).

Cookies are usually stored on your computer's hard drive in the folder of the respective browser. Most of them are safe (they are plain text files) and cannot affect the OS in any way. However, the technology that is supposed to make web surfing more convenient is increasingly being used to the detriment of the user – it is extremely difficult to control what information is collected about them by cookies. Also, no one can guarantee the complete security of the process of exchanging this data – cookies can be intercepted or stolen to track a person's previous online activities or to access their accounts. For this reason, website owners are obliged to warn about the use of cookies (for example, in accordance with the EU General Data Protection Regulation GDPR or the California Consumer Privacy Protection Act CCPA). However, unless you agree to at least the minimum set of cookies, you may not be able to browse the site fully. The massive adoption of these new privacy standards by multinational companies has been called an example of the «Brussels effect», a phenomenon where European laws and regulations are used as a reference because of their importance. A 2024 study showed that the GDPR reduced both the number of website page views by EU users and website revenue by 12%.

The user can only prohibit the browser from using certain cookies or delete them from time to time. For example, in Google Chrome *to block third-party cookies* click on the three dots in the upper right corner and go to «Settings» → «Privacy & Security» → «Third-party Cookies» and choose between «Allow third-party cookies», «Block third-party cookies in «Incognito mode»

to delete cookies – «Settings» → «Privacy and Security» → «Clear history» and select «Cookies and other site data» (solid) or «Settings» → «Privacy and Security» → «Third-party cookies» → «View all permissions and site data» (selective). Unfortunately, the browser does not currently support the option «Delete cookies and website data when closing all windows».

Online scanners such as 2gdpr, Cookie Compliance Audit Tool (Chrome plugin), CookieServe, CookieYes, Piwik, etc. allow you to pre-check the cookies of a website by clicking on the link to it. Protection against various forms of tracking with the use of cookies is provided by browser *anti-trackers* such as – Bitdefender's Anti-Tracker, Disconnect.me, Ghostery, Privacy Badger, Privacy Possum, etc., as well as ad *blocking plugins* such as – Adblockplus, AdBlocker Stands, Genius PRO, Popup Blocker, uBlock Origin.

The next step in the development of session and permanent cookies for data storage was the emergence of Web Storage (or Document Object Model Storage). This technology allows websites to store a significant amount of information (up to 10 MB) on the user's device without sending it to the server each time and access it through a special algorithm. Unlike cookies, which can be accessed both on the server and on the client side, web storage falls exclusively under the purview of client-side scripts. Similarly to cookies, Web Storage includes *Session storage* and *Local storage*. Regularly clearing them manually (see Deleting Cookies) or using the plugins ClearLocalStorage, Click&Clean, Local Storage, LocalStorage Manager, OneClick Cleaner, etc. will help to ensure privacy.

c) browser fingerprint is a unique identifier of the user's browser and OS configurations, which is formed on the basis of the data collected by various website tracking technologies. Traditional identification methods such as IP addresses and unique cookies are not used. A digital browser fingerprint is a 32-bit hexadecimal number such as 249821af43932314621f1246618ea8e1, which is obtained as a result of processing all data received from the browser. It allows you to track users on the Internet with an accuracy of up to 94%. Ideally, all machines have a different fingerprint value (divergence) and this value will never change (stability). In this case, it would be possible to uniquely identify every machine in the network without user consent.

Depending on the settings of the web resource, it is formed on the basis of a different number of parameters (on average, from 7 to 15, with a maximum of more than 40), including: *IP address*, *user-agent* (data about the browser, the installed OS and the device itself), *browser and device settings* (screen resolution, screen size, colour depth; browser language, other installed languages; date, time, font settings; availability of plug-ins and their characteristics), *browser graphics technologies* (particularly, display of

graphics and 3D images by Canvas and WebGL – based on them, tracking algorithms generate another unique fingerprint), *WebRTC* (a plug-in for streaming audio and video content that allows you to determine the user's real IP address, even if they use a VPN), *JavaScript i HTML5* (Document Object Model properties – how a web page is interpreted by the browser; LocalStorage and SessionStorage stored data; Do Not Track settings), *HTTP-headlines* (information for processing requests/responses between the browser and the server; for example, Accept-Language headers indicate the user's preferred language), *CSS* (how the browser interprets and processes page style language requests), *cookie and supercookie settings*, *behavioural factors* (typing speed, mouse movement patterns, search queries, browsing history), etc.

The collection of a browser fingerprint is currently considered acceptable and, unlike cookies, does not require formal user consent. A cookie is more like a tracking tool – as soon as it is on your computer, the website knows where you are and what you are doing. A browser fingerprint is more static and uses the data it receives to determine who you are, but it cannot follow you. At the same time, cookies can be blocked or deleted, while a digital fingerprint cannot (since most of the information it transmits is important for Internet surfing, it is almost impossible to disable it). It is used to prevent fraud (online banking, advertising fraud, blocking suspicious accounts), internal analytics and optimization of web content browsing, and to create a personality profile for marketing purposes (gender, age, marital status, level of financial wealth, interests, habits, or even personal data). The digital fingerprint can be stored and shared with affiliates to identify users when visiting affiliate websites.

You can find out the uniqueness of your browser's digital fingerprint using the resources of 2ip.ua, AmIUnique, Browserleaks, CoverYourTracks, Creepjs, Deviceinfo, ipleak.net, ip-score.com, Pixelscan, webkay.robinlinus, Whoer (RU), etc. The more you try to protect yourself from being taken down, the more unique you become and the easier it is to identify you (the so-called information entropy). A separate problem is the ability of a user to have multiple browsers on one device, let alone multiple virtual hosts. Changing browsers (or using several at the same time) no longer completely solves this problem due to the existence of Cross-Browser Fingerprinting. You can simply block the tracking of some browser fingerprints, and then websites will not see your digital profile. But such hiding becomes suspicious for tracking systems, which can affect the correct access to a particular resource.

Main *countermeasures*: 1) *keeping the browser as average* as possible (e.g., using default settings, keeping the language, time zone and country unchanged, not installing plug-ins, etc.); 2) *manually changing* or installing *plugins to hide certain components* of the digital fingerprint (e.g. AudioContext Fingerprint Defender, Canvas Fingerprint Defender, Fingerprint Spoofing, Font Fingerprint Defender, NoScript, Random User-Agent (Switcher), User-Agent Switcher, WebGL Fingerprint Defender,

WebRTC Network Limiter, WebRTC Protect). Indeed, changing at least one element of the fingerprint results in an automatic correction of the hash sum of the user's ID (formally, it becomes different). Although the identification accuracy in this case is reduced by only 0.3% and there is a risk that certain data may not match each other, for example, useragent does not match the browser kernel and its version, IP address does not match the one used for WebRTC, etc; 3) the *use of anti-detection browsers* (replacing the original digital fingerprint with a fake one), for example, AdsPower (free – up to 5 profiles), Dolphin-anty (free – up to 10 profiles), Ghost Browser (free – up to 4 profiles), Incogniton (free – up to 10 profiles), Mullvad (like Tor, provides all users with the same fake fingerprint), Switch Antidetect (free – up to 5 profiles, access via VPN), Undetectable (free – up to 10 profiles)), etc.; 4) *use of a Dedicated Server* a separate physical computer that does not transmit any information about the end user and his/her device. It is relatively expensive and requires professional configuration. Linux and Windows use different security software for scanning systems and networks to detect intrusions.

To neutralize certain vulnerabilities that can identify the user, you can use a Virtual Machine that emulates the operation of a computer or smartphone. This platform uses only the allocated capacity of the device on which it runs, otherwise it is like a separate physical device that allows you to install and run software, sometimes even incompatible with the current OS. Therefore, when surfing the Internet, it has a completely different digital browser fingerprint than a real device and a separate environment from it, which has a positive effect on protection against various cyber threats (for example, running potentially dangerous applications that can damage the OS or affect the operation of other programs; searching the DarkNet). Virtual machines are divided into 2 main categories: 1) system (hardware) virtual machines that provide a full emulation of the entire hardware platform and, accordingly, support the execution of the operating system and 2) application virtual machines that are designed to run only applications.

The virtual machine can be uninstalled and reinstalled at any time with the desired parameters, configured for specific search tasks, replicated, etc. At the same time, such a platform requires a certain amount of RAM, CPU time, and disc space, which can seriously slow down the entire system.

To *start working with it*, you need to: download the appropriate program, such as KVM, Microsoft Hyper-V ((built-in to Windows OS), QEMU, VirtualBox, VMware Workstation Player (free – runs one virtual machine on a Windows or Linux PC)), etc.; install it according to the documentation; download the ISO-image of the desired OS, create a virtual machine, assign resources to it, and start it; configure the guest OS, install the components required by the program, and perform search tasks.

So, which OS should you choose as a base for open source intelligence? The answer to this question depends on your needs and level of knowledge. In particular, there are a number of OSINT-oriented builds based on Linux – CSI Linux, Kali Linux, ParrotOS, Tails, Trace Labs OSINT VM, Tsurugi Linux, Whonix, that offer a wide range of tools right out of the box, ensure proper anonymization of the information search process, and allow you to create an isolated software environment when running through a virtual machine or from portable storage. They require certain skills to install and configure. The large number of specialized Linux distributions developed and maintained by various communities provides a wide range of software choices. So, the best choice is an OS that you know how to configure and feel comfortable working with.

A smartphone (or tablet) is now a whole ecosystem that includes messengers, social networks, and some other applications that are created exclusively for them. A mobile OS emulator is a tool for creating a virtual Android or iOS device on a computer and running the corresponding software shell. To imitate a real device, emulators either use their own virtualization implementation or work with off-the-shelf virtual machine options.

The most popular *Android emulators*, such as BlueStacks, KoPlayer, LDPlayer, MEmu, MuMu Player, NoxPlayer, XePlayer, allow you to choose a smartphone model, create its IMEI, mobile operator number, install applications from Google Play or apk files, and activate Root access (free versions have ads in the form of a block of recommended applications at the bottom of the main screen). Most products are free, but you need a Google account to get Root access.

Among the *iOS emulators* are Air iPhone, Appetize.io (browser-based application, free – one user, two active devices, session duration – 3 minutes, 30 minutes monthly), iPadian (free – limited number of applications in the App Store, advertising) and Xcode Simulator (part of Xcode, Apple's proprietary development environment).

Malware (Malicious Software) is a general term for a number of online cyber threats, including viruses, spyware, trojans, adware, worms, ransomware, etc. It can manifest itself in the form of code, script, active content, and other software.

Malware can get on your device through phishing, opening malicious attachments or emails, dangerous downloads, social engineering, removable media, or OS vulnerabilities. Software that can be classified as «malware» can be based on different technologies and have a completely different set of functions and

capabilities. Special anti-virus software is used to protect against threats.

The situation is more complicated with web resources that are in constant multi-user access for a huge number of connections around the world. *Signs of a website infection include*: a browser message warning about a dangerous resource when you go to a website; redirection to another website when you try to access the resource or follow internal links; additional content on the website, any ads or pop-ups; slow operation or unavailability of the resource, etc.

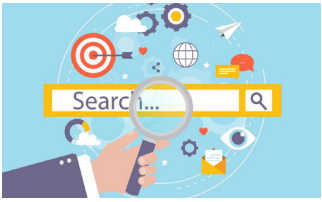
In the process of indexing websites, search engines check them for signs of infection, block dangerous content, and maintain and update blacklists of malicious resources. The only thing they cannot guarantee is the absence of malware in the files available for download on such pages. For example, you can find out the results of a website scan by Google's search service using Google Safe Browsing.

It is also worth using online services to *detect malware or suspicious links* on web resources, such as Astra Malware Scanner, phish.ly (email scanning), Quttera, ScamSearch.io (database of scammers and fraudulent websites), Sucuri SiteCheck, urlquery, urlscan.io, URLVoid, WhereGoes, *scanning of downloaded files* – Cuckoo Sandbox, InQuest, Intezer (requires signing up, free – for 2 weeks, then 10 scans of public files per month), Recorded Future Triage (requires signing up), Unpacme, Yomi or *universal solutions* – Any.Run (requires signing up), Hybrid Analysis, Joe Sandbox, malsub (GitHub), OPSWAT, Squarex (requires signing up), VirusTotal, as well as *checks against spam lists* – Blacklistalert.org, DNSChecker, DNSstuff, MultiRBL.valli.org, MXToolbox, RblHostingUkraine, SPAMHaus.org.

It is worth noting that it is practically impossible to provide 100% protection against malware using only the listed resources. In order to prevent infection and avoid negative consequences, it is advisable to use comprehensive solutions based on paid antivirus software.

Nowadays, many antivirus packages are not supplied in a «pure» form, but with many additional functions. These include, for example, the ability to back up important data, a smartphone or tablet security app, an advanced firewall, parental control functions, the ability to use it on multiple devices simultaneously, the ability to permanently destroy files and encrypt them, the ability to enable VPN, real-time web protection, email scanning, etc.

4. Universal Search Tools



A search engine is a website that allows searching for information on the Internet. The user enters a search query to start the search process and then receives a systematic list of links that are most relevant to the query. In common parlance, a search engine is a website that hosts the system's front-end. Information can be conveniently searched for by keyword or phrase (a set of words or phrases that most accurately reflect the essence of the information required).

Each search engine has its own algorithm, based on the following stages: of *scanning* (a special robot constantly searches for new web pages, adds them to the list of already known ones and saves the posted content); *indexing* (the uploaded pages are processed using various lexical and morphological algorithms, structuring and adding information to the search engine database); *sorting* (the program analyses the user's query, selects the pages that are most related to it from the database and displays the results in the form of a list of sources, sorted by the user's preferences).

For a long time, Google has been the most popular search engine (as of 2024, it accounts for more than 92% of all searches in the world), which is why the verb «to google» has become synonymous with searching for information on the Internet.

Features of Google search:

– the search engine reads the query from left to right, ignores the case of letters and punctuation marks, can conjugate words, and is limited to 32 words in length; the search is performed in the query language;

– the first words in the query have a greater impact on the relevance of the search results; by default, there is an invisible logical «and» between the words in the query;

– recognises text in documents (.pdf, .rtf, .docx, .xlsx, .pptx, etc.);

– the search result is personalised (depending on the user's location, device type, preferences, previous queries or websites viewed, advertising, etc.) and is not necessarily an exact match of the query (different case, number or synonym); non-personalised results are provided through anonymous browsing mode or by using the command google.com#q=google&pws=0;

– the search engine recommends a list of useful content formats – *Google Featured Snippets* (a block with short answers displayed in a separate window and located at the zero position of the search results);

– the list of sources may be affected by censorship due to violation of the rights of certain persons (copyright, right to be forgotten, etc.);

– Google cannot index information that can be accessed only by authorised users or after filling out certain forms, nor can it correctly retrieve data from video and audio content.

Google is a giant in the field of internet search, but it is far from a monopoly. While other search engines may not be as well-known, they still process millions of search queries daily. Therefore, when seeking specific information, it is always wise to leverage the capabilities of alternative search engines in addition to Google, especially when the latter does not yield the expected results. *A difference of just a few links can be crucial if it leads you to exactly what you are for so long looking for*:

• Microsoft Bing (market share ~ 3,3%) – in terms of functionality, Bing and Google are quite similar, but their search algorithms differ slightly. Bing tracks more interconnections between individual websites, provides better image and video search capabilities, and displays a variety of additional data in widgets. For convenience, it offers a search query history and collections (dedicated search projects for gathering diverse data with the option to continue them). It also integrates the BingAI chatbot (powered by OpenAI GPT-4) to enhance search capabilities and includes an interactive chat feature for creating text content, such as social media posts, emails, articles. As well as dynamically adjusting the amount of information displayed for each search result.

• Yandex (RU, market share ~ 1,5%) – focused on the Runet, it is blocked in Ukraine (accessible via VPN) and is under the control of Roskomnadzor, the regulator of the aggressor state. It enables effective searches for photos, videos, and content from popular social networks;

• Yahoo! (market share ~ 1,3%) – in 2009, the search engine was acquired by Microsoft, and since then, all search queries through Yahoo! are processed using the Bing system. Specializes in searches related to news, sports, and finance, though it remains relatively unknown to most users;

• Baidu (market share ~ 0,9%) – often referred to as the «Chinese Google», this platform supports only the Chinese language and is rarely used within the Ukrainian segment of the Internet;

• DuckDuckGo (market share ~ 0,6%) – An open-source search engine that, in addition to its own crawler uses the results from other search engines (including Yahoo! and Bing), providing more relevant outputs. It markets itself as the anti-Google, as it does not collect user data, does not personalize search results, anonymizes search history, and does not track cookies. Unlike Google, which personalizes results by region, DuckDuckGo allows users to easily search for information in other languages. It also filters search results by deliberately excluding links deemed «disinformation». DuckDuckGo is often the default search engine in privacy-focused browsers such as Tor Browser and Vivaldi.

Searching for information within the Ukrainian segment of the Internet can be enhanced through *domestic services* such as i.ua, META, Search, Yep and others.

There are advanced Google search commands (*Google Dorks*), that help refine and significantly narrow down search results, reducing the number of irrelevant outputs. They can also be used to identify vulnerabilities on web pages (the syntax for other search operators can be found in this list with Examples or through the GoogleHackingDatabase):

Request	*Examples*
Search by surname, name and patronymic	
All words (option)	G | новак іван петрович G | іван петрович новак G | novak ivan petrovych
Exact match	G |«Novak Ivan Petrovych»
One of the options	G |«Novak Ivan Petrovych» | «Novak I. P.»
All options	G |«Novak Ivan Petrovych» & «Novak I. P.» & «N.I.P.»
Grouping	G | новак (іван | иван) петрович G | (іван | иван | і.п. | и.п. | і. | и.) новак
Exclusion search	G | novak ivan petrovych -entrepreneur
By addition	G |novak ivan petrovych +attorney
Unknown letter, word, phrase	G |novak * petrovych
Range	G |novak ivan 2019..2024
Files of a specific type	G |novak ivan petrovych filetype:pdf (or .xls(x), .doc(x), .ppt(x), .rtf, .txt, .jpeg, .zip, .mp4 etc.)
In a location	G |novak ivan petrovych loc:Kharkiv
On forums (blogs)	G |novak ivan petrovych inurl:forum G |novak ivan petrovych inurl:blog
Profile search («nickname») on social networks	
On a specific site (domain)	G |(іван | иван) новак site:facebook.com (or on others – instagram.com, twitter.com, vk.com, ok.ru etc.)
By tags and hashtags	G |ivannovak@instagram (for X/Twitter, Facebook, Instagram) G |#Kharkovlawyers (for Instagram, VK, Facebook, Tumblr, TikTok)
Cache of deleted web page*	G |cache:https://www.rada.gov.ua/ G |cache:https://vk.com/id413593960 *since February 2024 it is no longer supported, though it might still work
Search for email by known nickname	
Word in the web page address (url)	G |inurl:novak_ivan site:mail.ru (allinurl: - all the words) (or on others – gmail.com, ukr.net, i.ua, meta.ua, yahoo.com, yandex.ru etc.)

Request	*Examples*
Search for a person's contact phone number	
	(новак іван петрович) & (мобильный | мобильник | мобила | моб. | сотовый | сотик | телефон | трубка | труба | мобільний | мобільник | mobile phone)
Search for information about a person in publications (news)	
Words in title	allintitle:resume novak ivan (intitle: - first word)
Words in text	allintext:attorney novak ivan (intext: - first word)
Similar website	related:https://traditionorder.info
Other words	novak AROUND (2) attorney (number of other words in brackets)
Synonyms	novak ivan ~attorney
Results before and after a certain date	novak ivan before:2022-02-24 novak ivan after:2022-02-24
Sites referencing url	link:novak.com

Other search engines also have similar Dorks. Additionally, multiple operators can be used in a single query (e.g., novak ivan site:*.ua filetype:pdf). Their functionality is duplicated in the intuitive interface of advanced search, as well as in the «All Filters» and «Tools» buttons located beneath the search bar in Google Chrome.

The **effectiveness of a search often depends primarily on the approach**:

– *creativity outweighs algorithms* – be creative when selecting keywords, envision the desired search result, combine variations of terms, add new ones, use synonyms and abbreviations, or put yourself in the perspective of the subject of interest etc.;

– recision can be enhanced by *using unique keywords* written in lowercase letters in the desired language and utilizing the «Find similar documents» function;

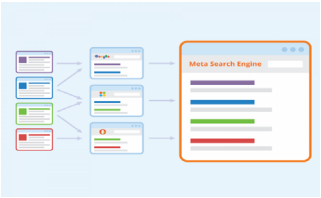
– review as *many search result pages as possible* – this increases the chances of finding less popular but more useful websites;

– *look for common file types on official websites* of local government bodies, municipal enterprises, educational institutions, healthcare facilities, preschools, and socio-political organizations. Such documents may contain relevant personal data, such as lists of students, housing or land allocation queues, lists of subsidized medications, payment documents, contracts, applications, complaints, decisions, and more.

Specialized services can be a useful addition – De Digger (searches publicly accessible files on Google Drives), FilePursuit (aggregates files from open web servers),

ISearchFrom (search from another region/device or in a different language), Keyword Tool (generates keywords based on a topic), Mark My Search (a plugin that highlights search query words on the page), Oldest search (displays the oldest web pages by date first), 2lingual (simultaneous search in two selected languages), as well as *online translators* – Bing.Translator, Deepl, Google and others.

Unfortunately, none of the existing search engines can independently cover all the constantly and dynamically growing resources of the Internet. Moreover, the search algorithms of different search engines for the documents they have already indexed differ to some extent. The need to expand search capabilities by aggregating the results of the best search engines in one place has led to the emergence of metasearch systems.



Such websites do not have their own databases or search indexes. When processing a query, they simultaneously poll several traditional search engines and return the results provided by them in a consolidated list without duplicate links, sometimes improving specific results. On one hand, this reduces effort and saves time, while on the other hand, the ranked results from multiple search engines can surpass the total output of metasearch systems in terms of quality. In other words, if there are many links on a topic, metasearch is unnecessary and may even be detrimental, as it mixes different ranking logics. However, if there are few links, metasearch can be useful precisely because it combines the limited results from different search engines.

The most common metasearch systems include: BizNar (which additionally provides search result of analytics and positions itself as a tool for working with the Deep Web – the part of the Internet inaccessible to regular search engines), BoardReader (searching information from online forums), Dogpile (presented as a search engine with the most comprehensive results and no ads), Excite, IntelligenceX (which allows users to choose search tools), Fagan Finder, Gibiru, Izito, MetaCrawler, metaGer, OSINT Helper, searX, Startpage, WebCrawler, Webmii, ZapMeta and others. At the same time, most of them have rather limited syntactic capabilities for building advanced search commands and working with Cyrillic characters.

The web archive, Internet Archive (The Wayback Machine) is a service for collecting and storing copies of websites (operating since 1996). It allows users to attempt to find data that has been changed over time (previous versions of web pages) or has become unavailable (deleted).

Similarly to the crawlers of search engines the Web Archive works by periodically visiting and storing publicly available content from websites

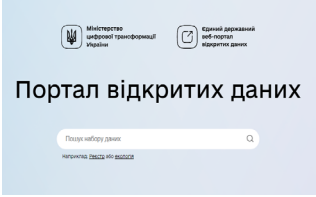
on its server according to a specific algorithm (content can be hidden from copying through passwords or indexing parameters of the resource) – HTML code, CSS styles and scripts; images, videos, music, documents, etc. When the Web Archive "visits" a site next time, it does not delete the previous copy but stores a new one. To perform a search, you need to enter the website's URL and choose the date for which you want to view the site's copy. There is also the option for users to *archive resources manually* (the "Save page now" option on the homepage).

The Archive.today service (launched in 2012) differs from the Wayback Machine in that it does not use bots and archives pages *only upon user request*. On the homepage, there are two prominent forms. The top red one allows users to archive, while the bottom gray one helps find a site among the saved ones. It ignores the standard access restriction for search robots through the robots.txt file, which allows it to index sites whose owners have prohibited archiving. It has several mirrors: archive.is, archive.li, archive.ph, archive.fo and others.

Additionally, it is worth mentioning services such as CachedView (search in the Google, Coral, and Internet Archive caches), Carbon Dating The Web (determines the creation date of a webpage), Quick Cache and Archive search (simultaneous search across 10 search engines and 24 web archives), suIP.biz (RU, search across multiple web archives), Web Archives (Chrome plugin) and Web-arhive.ru (RU).

If a page is missing from the web archive, it can be attempted to find via the search engine cache. This is the last visited and indexed version of the site by the search engine crawler – when the bot revisits, it overwrites the old version with a new one, deleting the old one; the update frequency ranges from one to four weeks. Unlike Bing and Yandex, Google removed the «cached» button from search results in February 2024, although the corresponding Dork continues to function.

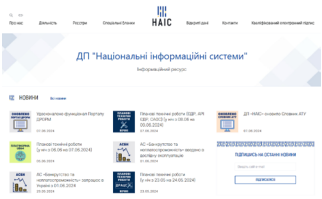
The Unified state web portal of open data in Ukraine is designed for free access and use of public information from government authorities, including its automated processing by electronic means. The declared goal is to ensure the transparent functioning of the government.



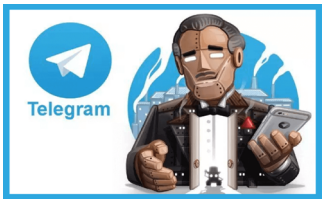
The informational and reference documentation (registers, directories, contracts, reports, decisions, orders, passports) is divided into 15 categories (e.g., state, economy and business, regional development, justice and judiciary), as well as by central and local authorities. The structure of the open data set includes descriptions of the composition (elements) of the data set, their format, parameters, and purpose (in formats such as .txt, .doc(x), .pdf, .xsd, .json, .csv, .zip, and others). Currently, the portal contains nearly *36,000 datasets* from subjects of authority and supports cross-search as well as search by groups or data controllers. The latter are

divided into central and local.

A related project, Diia.OpenData, contains step-by-step instructions for using datasets and the OpenDataToolkits toolset to further *enhance public oversight* of the activities of officials, government bodies, local authorities, judges, politicians, deputies, business entities, etc. This is achieved by i*dentifying signs of possible offenses* or *dishonesty* on their part, exposing corruption schemes in government procurement.



The main purpose of the state enterprise "National Information Systems" is the technical and technological support for the creation and maintenance of automated systems for Unified and State Registers operating under the orders of the Ministry of Justice of Ukraine, as well as other electronic databases (e.g., open data from registers). It ensures access to these systems and guarantees the preservation and protection of data. More detailed information about these functions will be discussed in Sections 7 and 8 of this publication.



Telegram bot (chatbot) is a robotic account in the Telegram messenger programmed to perform certain actions automatically, such as searching for useful information, selling products, creating content, etc. The application operates simply: it sends the user's message to a server as a command, processes the request, and returns a response through the bot, displaying it in the messenger. To use bots, you do not need to have any special knowledge. The utility automatically displays prompts and suggests certain commands. You only need to select the desired query. In addition, virtual assistants can display lists of command categories, which greatly simplifies the search.

Most Telegram bots designed to search for personal information about an individual (so-called «probiv») obtain data by processing open sources (search engines, social networks, forums, online advertisements, etc., utilizing the API of various sites), corporate database leaks (government agencies, telecom operators, banks, medical institutions, stores, insurance companies, delivery services, etc.), or their own resources (mobile applications for number identification, user survey results, advertising message mailings). In most cases, the services are paid (including via cryptocurrency), while free information is provided in a limited scope. Some applications include a referral system (a tool for attracting new users and earning certain bonuses) and offer the ability to create custom search bots (so-called mirrors).

The most popular bots for *universal search* include: BotoDetective (subscription-based), DataLeakAlert (RU, free – 1 week), EyeGodsBot (RU), HimeraSearch (RU), InfoBazaBot (free – basic functionality), LeakOSINT (RU, free

– referral system), LeakedInfoBot (RU, free – basic functionality),Nemezida2UA (RU, subscription-based), OpenDataUA (free – basic functionality), QuickOSINT (RU, free – 2 request), SmartSearchBot (RU, free – basic functionality), Unamer (RU), TSysBot (RU, subscription-based), Universal Search (RU, free – 10 queries, renewed every eight hours), UsersBox (RU, free – 7 request), Zernerda (RU, free – 2 request), Архангел (RU, subscription-based) and more.

Telegram bots are a convenient way to obtain necessary information, suggest a direction for search, or expand the set of initial data. However, *there are risks* associated with their use, such as the leakage of personal data, unauthorized access to the user's device, and malware infection through phishing links. Applications can make mistakes, provide incorrect or outdated data, or overlook important details. Therefore, it is always necessary to prioritize personal security, verify the provided information (use multiple bots), and not rely solely on it.

There is no universally accepted definition of artificial intelligence (AI) at present. However, AI in the broad sense refers to a range of technologies aimed at enhancing the ability of machines and intelligent systems (specifically computers) to perform tasks similar to human intelligence.

Once such systems are trained by humans, they can operate autonomously and continue learning from their results, gradually becoming more efficient. This is due, in part, to the fact that AI technologies identify patterns in data and then use these patterns for further predictions. In other words, AI systems are designed to learn from experience, recognize patterns, and make decisions based on input data. This is similar to how humans, through their neural networks, form consciousness and understanding of what is happening.

Some modern AI technologies include: machine learning, deep learning, neural networks, voice search, computer vision, pattern recognition, natural language generation (NLG), natural language processing (NLP) etc.

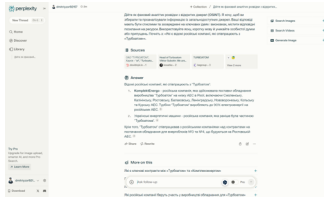
The U.S. intelligence community, recognizing the critical importance and growing scope of OSINT in its operations, has developed a new strategic initiative aimed at improving the collection, creation, and delivery of OSINT results by 2026. Central to this initiative is the application of AI and machine learning technologies, which are considered key to enhancing the processing of open-source data.

These technologies promise to increase the efficiency and accuracy of OSINT operations by automating the detection and analysis of relevant information from the vast amount of data available in open sources. However, the strategy also acknowledges challenges related

to ensuring the reliability and sufficiency of information, emphasizing the importance of developing robust data verification mechanisms.

Functions of search engines, which have become commonplace thanks to AI, include understanding queries in natural human language, voice search, reverse image search and geolocation, relevant ranking of search results, and translation of web pages.

Generative AI drew public attention in late 2022 with the launch of ChatGPT, the first widely accessible and easy-to-use chatbot based on this technology. Such tools can imitate human abilities to create texts, images, videos, music, software code, translations, and more. Today, millions of people use its capabilities in daily life, and the potential for adapting specific AI models to specialized fields of application seems almost limitless. However, there are also concerns about the potential misuse of generative AI, including cybercrime.



Given the content of the stages of the intelligence cycle, **multifunctional generative AI models**, such as ChatGPT, Claude AI, Google Gemini, Microsoft Copilot, Perplexity, You.com (with basic functionality available for free), are better suited for enhancing the effectiveness of OSINT research.

Their primary **uses** include:

– direct *search and data collection from websites* about objects of interest (news, profiles and publications on social media, forums, public registers and databases and more);

– *formulating keywords, ideas*, or a *list of resources* that can expand the search, *Google Dork queries* or X/Twitter searches to optimize the final output, as well as *command syntax* for *Github utilities*;

– *writing and/or optimizing script code for parsing* (automated data collection), *resource monitoring*, and other periodic tasks;

– technical *translation, transcription of video* and *audio content*, *processing unstructured data*, and *systematizing* it (e.g., mentions of people, organizations, events, or places, statistics; text, images, videos, links);

– *analyzing the collected information*, *preparing reports* (identifying trends, sentiments, threats, patterns, or connections, drawing conclusions and forecasts), and *visualizing it* (graphs, tables, charts, maps, diagrams).

Using the conversational mode, you can simply write queries to the chatbot in natural human language, just like you do in search engines, but this is unlikely to lead to anything useful. To get the most detailed and accurate answers, it is necessary to give neural networks detailed and clear tasks using prompts – a set of instructions for generating (forming) a specific result.

When we do not know how to formulate a promo, let AI do it for us. One of

the ways to create it is to use *constructors* – CHATGPT Prompt Generator, ChatGPT Prompt Generator, Chatgpt prompt generator, Free AI Prompt Generator, Gpt-Prompt (RU, for ChatGPT, YaGPT, GigaChat and Copilot), but it is better to create them *manually* following specific **rules**: *define the role* that the neural network should perform; *clearly formulate the query* (indicate exactly what you want to receive through step-by-step instructions, descriptions, lists of ideas, or comparisons of approaches to solving a task); *set mandatory conditions* (everything you write in [square brackets] must be followed by the neural network); if necessary, *divide the query into parts* (using quotes, sections, or numbered points so that the chatbot can work with large volumes of text or handle each part differently); *specify the format of the response* and *the style of presentation*.

For example, *"Act as an OSINT intelligence analyst. I want you to collect and analyze information from publicly available sources. Your answers should be concise, focused on key ideas and conclusions, and include appropriate references to resources. Use clear, brief language and avoid personal opinions or assumptions. Start with: "Find information about the cooperation of the company "ABCDCBA" (Kharkiv) with Russian companies after February 24, 2022".*

In other words, the more context you provide, the better the chatbot will be able to understand what you are looking for and respond more appropriately. Additional user's questions will help clarify previous answers and gather more information. If we do not know how to formulate a prompt, we can let the AI do it itself. It is important to verify the information generated (!) by the chatbot using alternative sources to confirm its accuracy.

At the end of July this year, OpenAI announced that it is working on its new AI search engine prototype called SearchGPT, which aims to provide users with quick and clear answers from relevant sources on the Internet. The company plans to later integrate its best features into ChatGPT. It could become the next stage of development for another interesting model, Globe Explorer. Additionally, there is a plugin store for ChatGPT – GPT Store.

5. Search by Photo and Video Content, Geolocation

It is not always possible to obtain the necessary information through text queries. Therefore, there is often a need to search for images and verify their authenticity, as they can intentionally or unintentionally mislead, spreading false information. Verification of photographs primarily involves determining the original source of the content, its author, location, date, and time of shooting, the circumstances of its appearance on the Internet, and whether there are any artificially modified elements.

Tools for determining:

1) the original source of photo content and its author (so-called reverse search – using a specific image as a query for search engines; programs scan the image, detect metadata, identify tags, and find indexed matches or related content)

• *general search engines* – Bing, Google Images, Yandex.Images;

• *Specialized search services* – Betaface (analyzing and comparing faces), Copyseeker, Faceagle, FactCheckTool (date and time of source indexing), Findclone (search for VKontakte, requires registration), Image Search, ImgOps, Lenso, Likelike AI ((search for Instagram), PhotOSINT (Chrome plugin, shows metadata), PhotoSherlock, PhotoTrackerLite (Chrome plugin), Reverse Image Search, RevEye Reverse Image Search (Chrome plugin), SauceNAO, Search4faces (RU, ; search among public figures, profile photos on VKontakte, avatars on VKontakte, Odnoklassniki, Tiktok, ClubHouse), Search by Image (Firefox plugin), Source search (RU, bot), TinEye, Who stole my pictures? (Chrome plugin). The resources Clearview.ai, FaceCheck.ID and PimEyes are paid but powerful search tools.

As a result of a reverse search, the search results will be a list of pages where the image you are looking for has been used, links to search functions for similar images, and possible related searches.

Reverse search may yield negative results due to social media account privacy settings, unindexed pages, or issues with the image itself (e.g., poor quality, bad angle, large time difference from the date of the shoot). Therefore, it is advisable to try searching the photo again after *technical processing* (e.g., scaling, color correction, isolating key fragments / faces, mirrored reflection, etc.):

• *comprehensive tools* – BeFunky, IMGonline, MMagic (GitHub), Online-Fotoshop (RU), OnlineVideoCutter, TinyWow.ImageTools;

• *image Quality Enhancement* – Bigjpg, Depix (GitHub, restores images hidden by blurring or pixelation), Ihancer, Image Enlarger (free – 10 credits per month, 1200x1200 pixels, up to 5 MB), Image Upscaler Online, Img.Upscaler (free – 10 credits per month, 2000x2000 pixels, up to 5 MB), LetsEnhance.io (free – 10 credits and watermark), MyHeritage (requires registration), Upscale.media (free – 3 images),

Waifu2x.net;

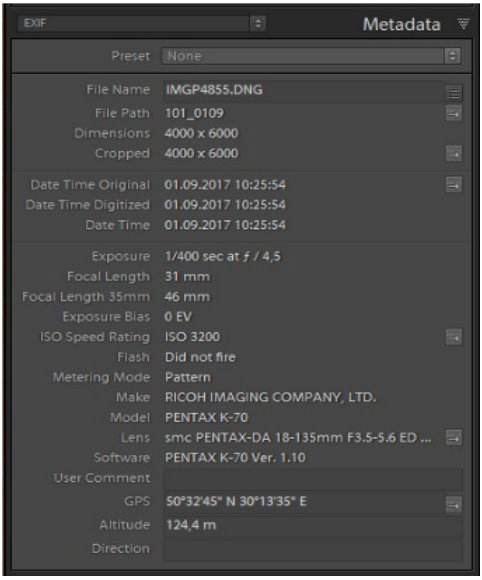
• *removal of unwanted objects* – Aiseesoft Free, BGbye, Background Remover, Cleanup.pictures (free – 720p resolution), Clipdrop, Inpaint (free – low resolution), Removebg, Watermark Remover.

The origin of the photo is typically identified by the earliest appearance date in the global network or by its higher resolution.

Verifying the origin of an image (investigating its «digital footprint» – who uploaded it and how) involves analyzing the source, its online history or associated resources/profiles, appearance time, interaction with other users, and previously/alternatively posted content. Furthermore, photos may be accompanied by descriptions, tags, comments, or specific text fragments that identify it – so it is important to extract potential keywords (e.g., acronyms, place names or descriptions, slang, etc.).

Pay attention to suspicious signs, such as a recently created account, a profile with few posts/followers/subscriptions, sudden changes in the geographic location of the author, or signs that they may be a bot (e.g., by comparing their followers with the ones they follow);

2) EXIF metadata (Exchangeable Image File Format) is additional information about a photograph stored at the beginning of the file, before the actual image data. It may include details such as the make and model of the camera, camera settings, date and time of the shot, coordinates, and the software used for processing the image etc. This information can be obtained through the "Properties" tab in the context menu of the file or by using software for viewing or editing photos/metadata, such as ExifTool, Exiv2, GIMP, GeoSetter, IrfanView), *online services* – CameraSummary, Exifdata, Fotor, Get IPTC, GroupDocs, IMGonline, Jimpl, Metadata2go, snapWONDERS (also works with video, free – 10 queries per day), ViewExifData or *browser plugins* – EXIF.tools (for Chrome), ExifViewer (for Chrome), xIFr (for Firefox).

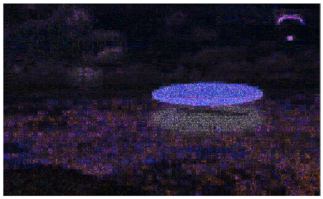


Similar information can also be obtained for video or PDF files, text documents, and websites (the <meta> tag in the head section of the HTML code). It is recommended to use several different sources to verify the accuracy of the obtained data.

At the same time, metadata is not a panacea – there is software available

in the public domain that allows extracting or modifying metadata from files. Social networks and messengers automatically remove EXIF data immediately after uploading or sending such a photo;

3) authenticity of the image can be verified through layered image analysis (photo forensics) using tools like Aperi'Solve, Digital Image Forensic Analyzer, Fake News Debunker (Chrome plugin, also works with video), Forensically, FotoForensics, Ghiro (Windows software), Image Edited?, Image Verification Assistant, JPEGsnoop (GitHub), Sherloq (GitHub). The main advantage of such services is that they work like a microscope, helping to see changes in the images that the human eye cannot see The most used *method is ELA* (Error Level Analysis), which analyzes compression artifacts in graphic files.



Typically, these artifacts are uniform across the entire image. If significant compression artifacts (such as bright, dark, or rainbow-colored areas) are found in specific elements, it is highly likely that these parts have been altered, indicating potential image editing in an image editor. This method also helps detect changes in brightness or contrast, such as the appearance of white spots in these areas.

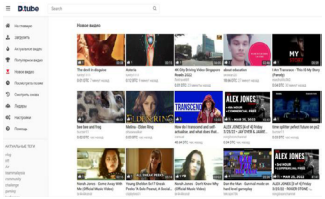
However, such services can be ineffective if the photo has been saved multiple times, especially with compression, as editing traces may disappear. Therefore, these tools may not be useful for analyzing images from social media, as their algorithms heavily compress images for uploading and remove EXIF data. But significant noise in ELA (blue and red stripes) is a sign that the photo has been re-saved multiple times.

The *Hidden Pixels method* will reveal hidden pixels, or transparent layers, that might have been used in image editing. These can help indirectly identify the apps used to edit the photo. For instance, Photoshop colors hidden pixels white, while Gimp and PicMonkey color them black.

Keep in mind that no single tool can definitively prove or disprove manipulation of a photograph. It is recommended to use multiple methods and resources for a more comprehensive analysis. The JPEG algorithm works quite cleanly on flat color areas and gradients, and much more wrong on sharp transitions.

In particular, determining the authenticity of an image will be helped by: *critical review of the image* (what looks strange – uniformity of lighting, unnatural shadows, mirror reflection, distortion at the edges of objects in the image, color changes, etc.); *researching the scene from different angles*; *determining changes in the landscape* (construction, combat actions, natural disasters), *content analysis of the image* (its magnification, color inversion to identify hard-to-notice details) together with

accompanying text, other photos taken earlier/later; *additional reverse search of the original image*; *checking the use of generative AI* – AI image Detector, AI or Not (free – 10 requests per month), AmIReal?, Content at Scale, Face Match, Fake Profile Detector (Chrome plugin), Illuminarty (free – basic functionality), Maybe's AI Art Detector, Hive AI Detector (Chrome plugin), а також Diffchecker (comparison of photos), PicTriev (gender and age of the person in the photo), Stolencamerafinder (determines the particular camera by serial number and searches the Internet for other photos taken by it).



Every day, users upload a huge amount of video content to the global network. However, its frame-by-frame indexing by search engines would require a significant amount of time and resources. That is why reverse video search, unlike images, has not become a basic functionality of these systems.

Bing, DuckDuckGo, Google Video, Yahoo Video Search, Yandex Video are limited to finding videos by keywords (title, description, participants' names, place, features, etc.) with the ability to filter results (by source, date, duration, volume) or by using Google Dorks (for example, site:youtube.com Ukrainian formula of peace). The same situation occurs in social networks (Facebook, Instagram, TikTok, VKontakte, etc.) – it is enough to enter a search query and go to the tab.

Therefore, the **content search and verification of videos** is very similar to working with images. When viewing the video, we take screenshots (for Windows OS the combination is Win+Shift+S, for Mac OS – Shift+Cmd+3), extract several unique fragments (beginning, key scenes), and perform reverse search to determine the original source. Then you need to filter the results to display only the videos Process automation tools – InVID WeVerify (Chrome plugin) and YouTubeDataViewer.

Undoubtedly, the biggest video hosting service is YouTube, where the largest volume of video materials is stored. The built-in search system on this video hosting is as efficient as Google's (not surprisingly, YouTube belongs to Google). A very important section of the service is live streaming (streams) прямі трансляції (Streams, стріми). Other popular free video hosting sites – Dailymotion, DTube, RuTube (RU), Dzen (RU), Video@Mail.Ru (RU).

Useful resources – Altoolskit (search among YouTube trends), Catchvideo.net (video download), Filmot (search in titles/subtitles), Hadzy (statistics of video comments), inPhrase (search for video and audio files by description), MW Metadata (view metadata), Return YouTube Comment Username (Chrome plugin, commentators' nicknames), Selectext (Chrome plugin, copy text from video in the browser), SnapSave

and SSyoutube (file download), TurboScribe (creates subtitles/annotations from video, free – 3 videos per day), YCF-Comment Finder (comments to the video), YouTube Channel Finder (free – 5 requests per day), YouTube location and Youtube-geofind (search for videos with geotags), YouTubeScreenshot (screenshots from video), YouTube search tool (search query constructor), YouTube Transcript (creates annotation if subtitles are available), YT1s.com (file download).

Important: The timestamp when uploading a file to YouTube is in Pacific Time (UTC-8, meaning 10 hours behind Kyiv).

When determining the authenticity of a video, *attention must be paid to details* (since software algorithms can be deceived): mismatched audio tracks, traces of editing («splicing»), video reflection (strange, «inverted» text), artificial enlargement/reduction of the image, adding new elements (time and date, bright logos), changing the color scheme to black and white, etc. *Countermeasures* – viewing EXIF, searching for other content that captures the same or similar event, which could potentially be passed off as the video under investigation; comparing frame references with satellite images and geolocation photos; analyzing the sound track (language, slang, names, events); checking weather conditions, frame-by-frame review in a video editor, etc.

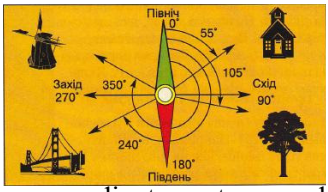
Tools for working with video – Anilyzer (frame-by-frame video review from YouTube), DFSpot-Deepfake-Recognition (GitHub, to determine if the video has been manipulated or generated synthetically), FlexClip (set of applications, free – 720p resolution, video length up to 10 minutes), Kapwing (online video editor, free – 720p resolution with watermarks), TinyWow.VideoTools (set of applications), Free Video Translator (bot, translation of YouTube videos into Ukrainian, annotation).

Geolocation analysis is the determination of the location of an object (stationary or moving) on a map, presented in the form of geographical coordinates, a postal address, or a movement route. **Determining the location** where a photograph (or video) was taken typically occurs **by**:

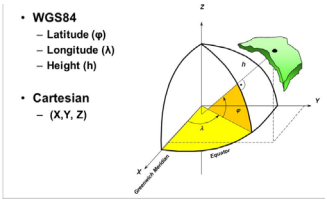
1) reviewing the presence of geospatial metadata (including GPS coordinates and their accuracy, altitude above sea level, time, speed, and azimuth of the GPS receiver's movement, azimuth of the image capture, and the type of geodetic system).

On Google Maps, Bing Maps, and Yandex Maps the search window first displays the *latitude* (from 0° to 90° north, from 0° to -90° south of the equator), followed by the *longitude* (from -180° to 0° west and from 0° to 180° east of Greenwich) in degrees (°) in decimal form (e.g., 10.35673009313803, -61.44516182050989). The service will point to the corresponding location on the surface of the planet. The coordinate format can be changed (e.g., to degrees, minutes and seconds) – Converting coordinates, Geo Coordinates Parser and Converter, Transform coordinates.

Coordinated Universal Time (UTC) approximately corresponds to the solar time at the Greenwich Meridian. Introduced in 1961, time zones around the globe are described as positive or negative offsets from UTC. However, UTC does not change in winter or summer, so in such countries there is a shift relative to UTC (Kyiv time is UTC+2 in winter, UTC+3 in summer).



Azimuth is the angle between the direction of north and the direction of an object, usually measured clockwise from the selected initial direction, and it can range from 0° to 360°.



A geodetic system is a coordinate system used to determine the exact location of an object on Earth. The standard for the global positioning system (GPS) at present is the 1984 World Geodetic System or WGS 84. WGS 84 determines coordinates relative to the Earth's center of mass, the error is less than 2 cm.

WGS84 is an improvement of the previous versions of the WGS72 and WGS66 systems. Like previous systems, it is related to the global ellipsoid, but it has refined dimensions and is oriented in such a way that its surface most accurately matches the physical surface of the Earth across the entire globe. The widely used system in Ukraine, Pulkovo-1942 (SK-42), employed the Krasovsky ellipsoid, which best matched the physical surface of the Earth only within the territory of the former USSR.

Resources for *geotagging photos* with metadata include – GeoImgr (free – 5 photos per day), geOSINT (GitHub), GpsPhoto (RU), Pic2Map, WhereIsThePicture. If there are multiple photos with GPS tags, placing them on a map in chronological order can provide a movement route of the object.

2) conducting reverse image search (or using a screenshot from a video) to establish potential matches with known objects or locations, including through *specialized services* such as EarthKit («Geoestimation», «Street View» and «Satellite» modes), Geolocation Estimation (AI landscape analysis to determine possible geolocation), GeoSpy, GVision (GitHub, free – 1000 request per month, requires Google Cloud Vision API), Landmark Recognition (recognition of natural and architectural landmarks), Picarta (AI analysis, free – 3 request per day), Wikinearby (popular places near specified coordinates).

A good practice is to remove irrelevant elements from the frame (such as people, animals, vehicles, furniture, etc.), pixelate or blur them to focus the search algorithms on the background – the landscape or interior. Also, analyze the name of the photo (or video), captions, and other accompanying information. If the object in the image can be described in words (for example, hotel, mountain, church and more), try to search for them in the language of the country of the potential location;

3) detailed examination of the image (focusing on visual landmarks) – view from the window, features of surrounding buildings, their addresses, notable landmarks, outdoor advertisements, street signs and streetlights, traffic signs and

road markings, power lines, road configuration, writings or drawings, license plates, transport routes, characteristic terrain, weather conditions, sun shadows, constellations, interior items, clothing details, etc. – Animal.toolpie (animal identification), Camopedia (military uniforms), Geohints and GeoTips (geospatial archive of objects and their features for different countries), Geonames (global geographical database), Logo.toolpie (logo identification), Plant.toolpie та Pl@ntNet (plant identification), PlatesMania (RU, license plates of countries worldwide), Plate recognizer (license plate recognition, free – 2500 requests per month), Vehicle AI (vehicle identification from photos), Worldlicenseplates (license plates orecognition);

4) visual confirmation of the geolocation follows the principle of moving from global to local landmarks (country → region → settlement → street → building or specific area):

• *satellite maps* (free – relatively low resolution of the image) ArcGIS, Bing Maps, CopernicusBrowser, EO Browser, ImageHunter (archive of images of the selected area), Google Earth (in the PC version, Google Earth Pro has more advanced functionality and an image archive), EOS Land Viewer, Google Maps, HereWeGo, OpenAerialMap (relatively few images), Satellites.pro (includes maps from Google, Yandex, OpenStreet, ESRI and Apple), Soar, World Imagery, World Imagery Wayback (archive of images since 2014), Yandex.Maps (RU); *fire maps* (places hit during combat actions) – Greenpeace, Firms.NASA, Pogodnik.com, SaveEcoBot, Worldview (archive since 2021);

• *street view, webcams, and dashcam photos* – Google Street View (for a start drag the person icon on the panel at the bottom right to the right place), Yandex.Panorama (RU); Instant Street View and ShowMyStreet (street view in Google Street View); webcams – EarthCam, Insecam (RU), Opentopia, Pictimo, RailWebcams (webcams related to railways), Shodan (search for devices, including webcams connected to the internet), Skyline, Surveillance under Surveillance, Windy Webcam, WorldCam, WorldCams, World-Cam (RU), YouWebCams (RU), Google Custom Search Engine (search for a catalog of webcam services); KartaView and Mapillary (street/dashcam images), WindowSwap (views from around the world, paid location selection), 360cities (panoramas);

• *specialized maps* – Bellingcat OpenStreetMap search (search for objects by categories in a given area), CellMapper (mobile operator base stations), DualMap (simultaneous display of a location via street view, regular map, and 3D projection) GpsJam (GPS signal interference based on aircraft navigation system reports), MapCompare (simultaneous work with multiple maps from over 250 available in the application), OpenCellid (cellular network base stations), OpenInfraMap (lines, power transmission, telecommunications, solar, oil, gas, and water infrastructure worldwide), OpenStreetMap (building boundaries of various purposes, transport objects); Peakery, PeakFinder, PeakVisor (mountain landscape identification), Wikimapia (RU, an interactive map with the designation and description of geographical objects), 2GIS (RU, requires VPN, city maps with reference information), Visicom (maps of Ukrainian cities with additional data);

• *Russian military aggression* – Bellingcat's map of Ukraine, DeepStateMap, Eyes on Russia Map, Geoconfirmed, LiveUAMap, [WarArchive] War Map, map of Russian military objects in Crimea;

• *auxiliary services* – Earth Engine Data Catalog (download Earth Engine catalogs), Geofabrik (free map data for OpenStreetMap), GEOINTsearch (search for Google Maps coordinates links from X/Twitter, Reddit, and 4Chan), MapSwitcher (plugin for Chrome, switches electronic maps, coordinate conversions), Old maps online (archival maps), Pastvu (RU, archival photos of settlements), QGIS (open-source geographic information system), SearchOnMap (RU, bot, search for objects on a map – buildings, paths, bodies of water, etc., by their parameters); *measurements* – CalcMaps (RU), Free Map Tools, Grid Reference Finder, MapChecking (approximately determines the number of people at a given location), Maps.ie;

5) determining approximate date and time (weather conditions) of the shot (chronolocation):

• *3D maps* – F4map (free – buildings and their shadows in real-time), OSM Buildings, Skydb (database of skyscrapers and high-rise buildings);

• *Sun/Moon shadows* (display of the Sun/Moon movement trajectory on a specific day at any point in the world for further comparison of shadows in the photo/video with service data for the location of the shot) – 3D Sun Path, Gaisma (sunrise and sunset times worldwide), MoonCalc, Online Protractor (protractor), Scale fixereng (measuring objects in a photo), ShadeMap, ShadowCalculator, Shadowmap (free – select location, only for the current day), SunCalc;

• *weather archive* (temperature, cloud cover, precipitation, solar angle of incidence, etc.) – Wolfram Alpha ((English-language resource, example query – weather kharkiv 10 april 2024), Ventusky, Windy, Zoom.earth (interactive map).

6. Socially-oriented Platforms

Social network is a website or application where users can create personal profiles, virtually exchange information, maintain contacts with other community members, and form new acquaintances. People want to share experiences, emotions, views, news, products, or simply communicate because they depend on each other for approval, recognition, and socialization, as well as for conducting business. The nature and nomenclature of links may vary depending on the system.

At the same time, social networks store large volumes of data about their users, including their public posts, photos, videos, social circles, and activity history, making these platforms ideal for OSINT research. Processing this data requires the use of virtual accounts. However, the owners of these platforms are constantly updating the functionality to protect the privacy of authorized users, which sometimes complicates the work of OSINT researchers.

The presence of a person's profile on social networks can be determined through:

• ***search engines and their operators*** – for example, site:linkedin.com/in «NLU» (this will bring up a list of LinkedIn accounts related to the Yaroslav Mudryi National Law University), site:(https://www.facebook.com/ | https://vk.com/) «Novak Ivan» will search for the user Ivan Novak among profiles on Facebook and VKontakte), etc. However, much of the information on social networks is stored in the Deep Web and therefore is not well indexed, making search engines not always an effective tool;

• ***specialized services*** (basic search by nickname) – Alfred (GitHub), Blackbird, Check Usernames, DetectDee (GitHub, search by email and phone), EagleEye (GitHub, search by photo), Google Social Search, Enola (GitHub), Holehe (GitHub, search by email), Maigret (GitHub), Maryam (GitHub, social_nets module), Mr.Holmes (GitHub), NameCheckup and NaMint (used in reverse – looking for platforms where a nickname is unavailable, indicating that the person may have an account there), OSINT Tools (search by email and phone), Predictasearch (search by email or phone, free – basic functionality), Seon (search by email or phone), Sherlock (GitHub), Sherlockeye, Snoop (GitHub), Social Analyzer (GitHub), Social Mapper (GitHub, search by name, email, phone number, photo), SocialRecon (GitHub), Social Searcher (free – 100 queries per day), Username Lookup, Usersearch.org (free – basic functionality), WhatsMyNameWeb;

• *reverse image search*;

• ***Telegram bots*** – for universal search, as well as *niche-specific ones* – GetFB (RU, search by phone number), Maigret OSINT (RU, search by nickname);

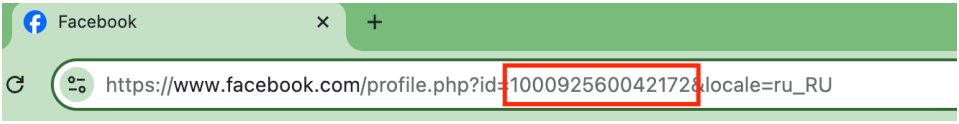
• ***search capabilities of social networks*** – *by nickname or full name* (including different spelling variations), *by phone number* (by adding the number to the contacts list and the person to friends), *by known profile on another social network* (through the friends list, resume), *through subscriptions* (common acquaintances), *by photo tags* (participation in common events); *by hashtags* (profession, occupation or interests, event or places of study, work, leisure, etc.; third-party resources – Kribrum.io (RU, requires registration and VPN), Quick hashtags and keywords search, Storyful Multisearch (Chrome plugin)) and *geotags* (additionally – geOSINT (GitHub, searches social networks for photos with geotags and maps them), OSINT Geolocation Databases Search (search through Bellingcat, Cen4InfoRes databases, etc.), Social Geo Lens, xHunt (free – 7 days, requires registration)).

Very often, a person may use variations of their full name or initials as a nickname, adding their date (year) of birth, the name of a favorite literary or historical character, information about the person (workplace or profession, worldview/psychology, preferences), or «so no one will guess» – a word written backwards, Cyrillic in the English layout, Latin letters, etc.;

• ***commercial solutions*** – Artellence, Tangles, etc.

Based on the search results, it is important to make sure that the page you find really belongs to the object of interest. A user profile is a visual display of personal data related to a specific person and characterizing his or her unique environment on a social network. It is a digital representation of a person's personality. However, there is a risk that false or incomplete data was entered during authorization.

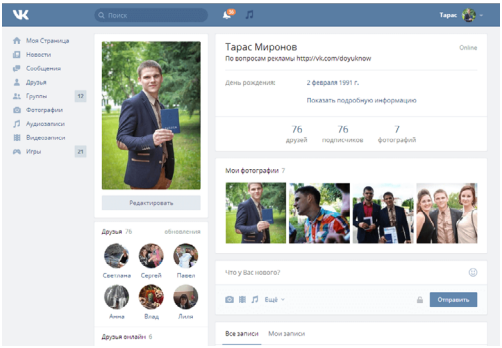
User ID – a unique, unchangeable number assigned to an account, community, public page, or group upon registration in a specific social network (each network has its own rules for creating IDs). This identifier helps to unambiguously authorize the account owner, regardless of their last name, first name, or nickname. User ID stores in the internal system of the site personal data of the user – name, phone number, email and others, as well as the history of his interaction with the application or the network's site – what orders he made, what mailings he received and others. For example, for Facebook, it is a 15-digit number following «id=» in the URL-адреси (Uniform Resource Locator) of the user's page.



The user ID and username are part of the public profile. It is also possible to change this identifier to a name in Latin letters, which is located after the domain name in the browser address bar (https://facebook.com/yourname). In this case, to determine the ID you can search the HTML-code of the page (by calling up the context menu with the right mouse button or keyboard shortcut) or use the functionality of the following specialised resources.

By ID, any other person can find the corresponding profile and view all publicly available information (the user can choose which information to make public in the security settings).

To learn more about a person, the following **indicators** can be helpful: profile (avatar, status, personal data); friend list; news feed, posts/ resharing, publications; frequency of material posting; photos, music, videos, locations; participation in groups, communities, subscriptions; maintaining personal blogs, including professional, socio-political, etc.



For this purpose, it is useful to apply both *universal tools* for most social networks – such as Comment Picker, OnePlus OSINT Toolkit, Shreateh Social Media Tools and *specific resources* – 4K Video Downloader (video downloading), CrowdTangle (Chrome plugin, shows who shared the link and how often), ExportComments (export comments on posts, free – 100 comments per day), NetSocOSINT (GitHub, gathers information from Instagram, TikTok, X/Twitter, Twitch, Telegram, GitHub accounts), Phantom Buster (gathers data from accounts, free – 14 days, limited functionality), Popsters (RU, account analytics, requires VPN, free – 7 days, 10 downloads), SaveFrom.net (video downloading), Social Blade (account statistics), Social Searcher (monitoring mentions in social networks), Tagdef (helps with hashtag meanings), Tonetizer (post tone analyzer, Ukrainian language not available), VideoDownloader (video downloading), Video DownloadHelper (Chrome plugin, video downloading), VideoGrabber, *as well as niche-specific tools:*

Facebook:

- *ID* – Codeofaninja, Find Facebook ID, Lookup-id.com, Randomtools.io;
- *alternative to internal search* – Facebook Hashtag Search, FacebookMatrix, Facebook Search, Facebook Search CSE, Graph.tips, IntelligenceX, IntelTechniques, Search is Back, Socmint Tool, SowSearch, WhoPostedWhat;
- *profile analysis* – StalkFace;
- *download content* – DumpItBlue+ (plugin for Chrome), Fdown.net;
- *other* – Facebook Data Breach Checker (check if the phone is used for registration), Facebook Live Map (search live streams), Facebook Recover Lookup (recover a profile).

Instagram:

- *ID* – Codeofaninja, Find Instagram User ID;
- *alternative to internal search* – Aware Online, Instagram Explorer (search a photo), IntelTechniques;

- *profile analysis* – Instagram-scraper (GitHub), Modash, Not Just Analytics, Osintgram (GitHub), SolG (GitHub), Toutatis (GitHub, receives data from the profile – email, phone, etc. via API), Unseen (GitHub), Webstagram;
- *view profile and download content* – DownloadGram, Dumpoir, ExportGram, GreatFon, iGram, Imginn, IMGinn.io, Inflact, Instafollowers, INDownloader, InstaFreeView, Instaloader (GitHub), InstaNavigation, InstaDP, Pixwox, Picuki, PokoInsta, SaveInsta, Snapinsta.app, StoriesIG, StorySaver.net.

LinkedIn:

- *alternative to internal search* – CrossLinked (GitHub, data on employees), CSE, Free people search, IntelligenceX, IntelTechniques, LinkedIn Boolean Search Tool (free – 7 days), LinkedIn Guest Browser (plugin for Firefox, view accounts without registration), LinkedInt (GitHub), Programmable Search Engine, Recruit'em, Recruitment Geek, RocketReach (requires registration);
- *profile analysis* – InSpy (GitHub);
- *download content* – Linkedin Video Downloader;
- *other* – LinkedIn Overlay Remover (plugin for Firefox; removes the overlay that that is displayed on top of a Linkedin profile).

X (Twitter):

- *ID* – Codeofaninja, Find Twitter ID, TweeterID;
- *alternative to internal search* – Aware Online, FollowerWonk, Free people search, IntelligenceX, IntelTechniques, Network Tool (ways to distribute tweets), One Million Tweet Map (tweets with geotags), Socialbearing (analytics), Synapsint, Tweet Archiver (Chrome plugin), Twint (GitHub, scraper), Twitter Advanced Search, Twitter List search, Twitter search tool, twXplorer;
- *profile analysis* – Foller, Lolarchiver (profile history), memory.lol (historical data), Tinfoleak, Twitonomy, TwitterAudit (authenticity of followers);
- *download content* – Download Twitter Data, TwitterVideoDownloader;
- *other* – Botometre (bot account detection), Deleted Tweet Finder (search for deleted tweets), SimpleScraper OSINT (monitoring tweets with coordinates), Wayback Tweets (GitHub, archived tweets), Wayback Tweets (archived tweets).

TikTok:

- *ID* – Find Tiktok ID;
- *alternative to internal search* – Aware Online, TikTok Quick Search, UrleBird, Vidnice;
- *profile analysis* – Exolyt (free – basic functionality), MaveKite (free – basic functionality), TikTok hashtag analysis (GitHub, hashtag analysis);
- *download content* – Ssstik, SnapTik, Tiker, TikTok Scraper (GitHub), TikTok Downloader, Tiktok Video Downloader, Ttdown, TTSave;
- *other* – TikTok-Timestamp (date and time of video upload).

VKontakte (RU):

- *ID* – ForVk (RU), Prozavr (RU);
- *alternative to internal search* – BigBookName (RU), Custom search engine Google, Photo-Map (RU, search for posts with geotags), Vk.barkov.net (RU), VK.watch (RU, additional – search by photo);
- *profile analysis and its connections* – 220vk (RU), FindNameVk (RU, bot), InfoApp (RU, requires a social network profile), Social Graph Bot (RU, bot, graphical analysis of pages), UseVk (RU), VKCity4Me (RU), VKUserInfo (RU, bot);
- *other* – ForVk, Nebaz та Vkdia (RU, user tracking), Regvk (RU, account registration date), VKHistoryRobot (RU, bot, profile archive).

Odnoklasniki (RU):

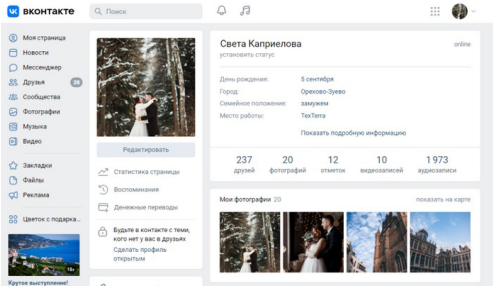
- *alternative to internal search* – poisk-cheloveka (RU), Vk.barkov.net (RU, additionally – profile analysis).

Discord:

- *ID* – Discord.name, Disserv (GitHub), Lookupguru, Unofficial Discord Lookup;
- *alternative to internal search (servers and bots)* – Disboard, Discordbots.gg, Discordbotlist, Discord center, Discord discadia, Discord.me, Discord official server search, Top.gg;
- *other* – Discord Client Encyclopedia (GitHub, a set of third-party clients and mods).

If reliable, the information obtained with the help of these tools may indicate: a person's psychological portrait, psychotype, character, openness; values, moral attitudes, level of culture and upbringing, perception of the world, motivation, priorities; hobbies, hobbies and interests, leisure, rhythm of life, social circle; emotional state, mood; level of aggression, conflict; public activity, civic position, political views, attitude to certain events or people; professional skills, achievements; reputation; likes/ antipathies, preferences and tastes, aspirations and ideals etc.

When looking for a person on a social network, **you need to consider all their incoming and outgoing interactions** – mentions of them in posts by relatives or acquaintances, organizers, the presence of joint photos, captions or comments, locations, etc. If a person is silent or has even closed their account from outsiders, their relatives or friends can be much more «open», giving the person away simply by the mere fact of their presence. For example, to find a profile, we search LinkedIn/ Facebook for their colleagues with unique names (preferably women –

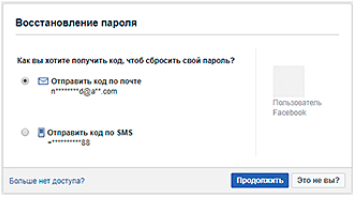


they often do not close their accounts and do not hide their real names), and then his or her own profile – in subscriptions/marks, comments, corporate photos, etc.

Photos on social media and descriptions under them can tell you about the places a person has visited (work/leisure, video tours of an office or apartment, travel, window views, group photos, elevator photos, geolocation marks, etc.), and friend lists will help you outline their circle of close contacts. Therefore, to find *photos of a person or their potential acquaintances*, find a place where they may frequent (school/work, beauty salon, gym, car service, entertainment venues, etc.) and browse the photos posted there, as well as the profiles of their authors.

People often use the same photo as their *profile avatar* for different social networks – to find accounts, use the reverse image search or Dorks (for example, ivan novak imagesize:170x170 site:http://facebook.com, photo and video sizes for different social networks). If necessary, enhance the image or determine its authenticity. It is better to find an abandoned account that has not been active for a long time and borrow a couple of photos from there. Put one of them in your profile, and put the rest on the page for plausibility.

Certain data can be obtained through the *«Restore Accounts» function* of the social network. If you enter a known email address or phone number, you will receive the message «No search results», which means that the address is unknown, for example, in VKontakte. However, when the address is known, you will be asked to change the password and parts of the backup mailbox address or number will be displayed, which can help in further searches (for example, to validate other user's accounts). Works in X/Twitter, Facebook, Instagram, VKontakte, Odnoklassniki.



Given the purpose of creating a virtual account, it may be necessary to further fill and promote it, at least at the level of activity of an average user (post something on your wall every few days, add a neutral comment under a post in a group, or like a post); occasionally join thematic groups depending on their stated interests or location, subscribe to relevant publics); having an optimal number of friends and/or followers (send friend requests to members of groups they have joined); sharing targeted, consistent, interconnected and uniquely related content (a message, a thematic picture, a meme or a repost of a post). Services for generating fake correspondence include – FakeChatMaker, FakeDetails, Pranx, Simitator, Zeoob or the use of AI capabilities.

The best practice is to create a virtual account in a language that the user is fluent in, as translated texts can immediately catch the eye and cause unnecessary doubts (alternatively, limit textual information as much as possible or use the help

of specialists). The more data you add to your profile, the easier it is to verify its authenticity – it should be based on real prototypes. An impersonal image (object, abstraction, character, or drawing) can be used as an avatar, although we subconsciously trust the interlocutor with a photo of a real person more. Sometimes it is helpful to use stock images and crop the photo so that any previously stored data is deleted before uploading, as social media platforms have algorithms that can detect the use of stock images and flag your account.

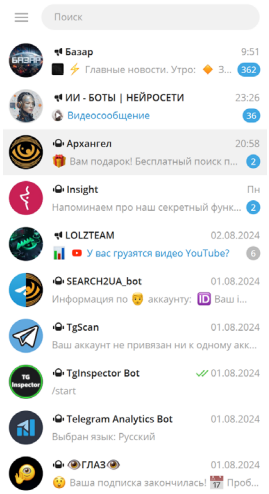
Social networks are often grouped with instant messengers: both provide a means of communication for many people. The main difference is that social networks provide access to mass communication. For example, your post on Facebook will be seen by all your friends at once, and maybe even by other six people at once. Whereas messengers allow people to communicate one-on-one or create small communities, such as a family chat or a chat for colleagues. The line between these concepts is quite blurred: many social networks have built-in messengers, and some messengers can publish posts to the public. Messengers usually provide end-to-end encrypted chats, video calls, VoIP, file sharing, and some other features.

Telegram (RU) has long been more than just a messenger - for many people, it is the main source of news, useful or entertaining content. It allows you to communicate with other users, create groups, channels, bots, and much more. The most common way to find out if a person is present in Telegram (if you have the app installed and their phone number) is to enter https://t.me/+XXXXXXXXXXX in the browser address bar and, if confirmed, go to the corresponding chat and view it. Alternative options are to add this number to the phone book and analyze the result; telegram-phone-number-checker (GitHub).

Telegram ID is a set of numbers that identifies a profile (username) in the messenger. It does not depend on the username, phone number, photo or other data, and cannot be changed or deleted. The standard functionality of Telegram does not provide a function that allows you to find out your own/other ID. For this purpose, the capabilities of bots are mainly used – GetMyID та UserInfoBot (RU, user ID, current chat ID, ID of the sender of the message in a public group or chat ID of the forwarded message from such a group), usinfobot, Userbox , Telerecon (complex OSINT research), bots for universal search.

It should be remembered that when you change your phone number, you do not need to create a new account in Telegram, as this messenger has the function of transferring your account to another number. At the same time, the message history, contacts, media, and other data associated with the account are saved. You can share photos, videos, and files of any type. For photos, the Internet search

function is available. The file size for sending is limited to 2 GB.



Channels and groups on a wide variety of topics attract huge audiences. *A channel* is like a thematic blog or page on a social network, to which you can subscribe and follow its life, information updates, view publications and respond to them, but in no way influence its activities. While a Telegram *group* is a community where you can post content yourself, participate in discussions, reply to participants, etc. Sometimes such communities have administrators who monitor violations of the group's rules. Is it possible to turn a group into a Telegram channel? You cannot – they are different entities. You can only create a new channel. A channel cannot be transformed into a group either.

Channels and groups have a similar feature – varieties (public and private). For example, *public* channels are available to all users of the are available to all users of the messenger through its general search through its general search. *Private* channels cannot be found in this way. The link for them is generated by the system, and only by such an invitation can users go to the channel and subscribe. But these types of group communication also differ in terms of administration, functionality, and other parameters.

Telegram channels have no restrictions on the number of subscribers. Only the administrator can send messages in them. For participants to communicate with each other, the administrator should add the ability to comment on posts. Communication in the comments takes place as a separate branch of the conversation that belongs to a specific message. Channel subscribers can also react to posts (just like liking them on social media, but with a slightly wider list of emojis for reactions). The moderator's functions are to remove participants who have violated the rules and impose permanent or temporary restrictions. Channel members can view the history of posts from the very beginning – new subscribers can learn about previous posts and even download them. An important element of subscribing to the channel is the privacy of the participants - we can only see the total number of them and do not have access to their profiles (personal data and phone numbers, only avatar and nickname). This option is available only in comments (as this is a separate group created specifically for communication).

When you join a group, in some cases you cannot view the post history due to administrator restrictions. However, in most cases, new users can read archived messages and posts. At the beginning of the messenger's development, no more than 200 people could join the group, and now the number of supergroups is limited to 200,000 participants. Group activity can be studied

only by those users who are in the group or leave comments. And in channels, the administrator can find out the number of views of each post. In the group settings, you can disallow reactions, certain types of content, or the creation of polls. And if you link a group to a channel, you can switch between them and have separate discussions under each post.

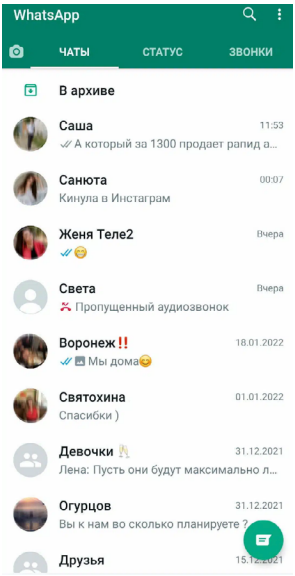
Groups on Telegram are a way of communication, they allow you to unite a large number of people in one dialog. Channels are intended to receive information from a specific source without the ability to directly influence it. Therefore, we will distinguish *tools for*:

• ***search for channels, groups and bots*** – DirectoryTG, Lyzem, Telegogo Google CSE (public messages), Telegram Channels, Telegram Channels Search, TelegramDB, TelegramGroup, Telemetr.io (free – basic functionality, requires registration, displays deleted posts in the «Posts» tab), Telemetr.me (RU), TeleScan (RU, bot, searches for groups, downloads messages), Telemetry (searches for messages, free – 5 requests per day with 25 results in the search results), TeleSINT (RU, bot, searches for groups in which the user is a member), TGInspector (RU, bot, searches for groups, downloads messages), TGScan, TGStat (RU), xTea;

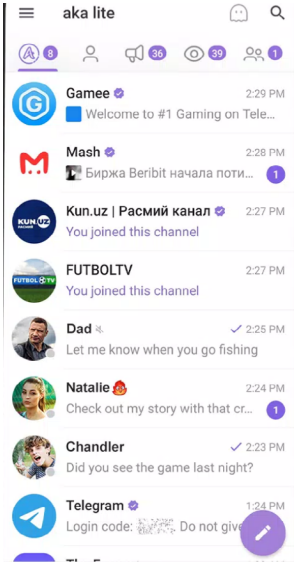
• ***research of user/group/channel profile*** – CCTV (GitHub, location tracking via API), CommentGram (search for comments), FunStat (RU, bot, various statistics), Geogramint (GitHub, search for users and groups that have activated the «Nearby» function via API; by default, it is disabled), informer (GitHub, information about channels, groups and users), IntelligenceX (search and analysis of Telegram data), Insight (RU, bot, user interests based on their activity in groups), LinkGrabber (collects links placed on a web page – friends, authors of a comment or a like), Save Telegram Chat History (GitHub, Chrome plugin, saving chat messages), Telegra.ph (RU, Telegram search engine), Telegram Message Analyzer (GitHub, analytics of the chat history saved in a .html file of chat history), Telegram Nearby Map (GitHub, determines the location of users nearby), Telegram Scraper (GitHub, information about group members), Telegram Sender (Chrome plugin, collects nicknames of group members), Telegram Tracker (GitHub, generates json files with information about Telegram chats and posts, requires API), Telepathy-Community (GitHub, allows you to archive Telegram chats, including answers, media content, comments and reactions, collect lists of participants, search for users by a given location, analyze the most popular messages in the chat, map forwarded messages, etc.), TgDev (RU, search for Telegram channel posts), TeleTracker (channel research), TlgGeoEarthBot (bot, displays active Telegram accounts with geolocation enabled around a given point; free – 3 requests per day), Tosint (research of bots associated with a Telegram channel).

In Telegram, the end-to-end encryption feature does not work until you create a secret chat with members of your group. In contrast, WhatsApp provides end-to-end encryption for chats and calls on its platform, which imposes certain limitations on the functionality of the *search tool*:

• ***checking the registration/connection status*** – WATools.IO (free – 8 hours to track a person's status, notify them of their WhatsApp use, monitor their activity, and study the likelihood of a chat between two people), similar functionality in Chatwatch (requires registration, free – 3 days), Whapi (additionally – Automatic warm-up of WhatsApp accounts, Chat Link Generator, Products, Profile picture, QR Code Generator, Text Formatter; free – 5 conversations per month, 150 messages and 30 requests per day, 1000 ARI calls per month per month), Whatsapp Mobile Tools, WhatsappMonitor (GitHub, activity monitoring), WhatsApp-Monitor (GitHub, tracker, notifications), WhatsApp OSINT (additional – user data, time zone in which it is located), WhatsApp OSINT Tool (GitHub, session duration);



• ***other*** – Email2WhatsApp (GitHub, searches for numbers registered in WhatsApp by email), Fake WhatsApp Chat Generator (fake chat), Whatsapp-GroupContacts-Scraper (GitHub, вdownload contacts from WhatsApp group chats), WhatsFoto (GitHub, Chrome plugin, download profile photo), WhatScraper (GitHub, download information about group members).



Viber like any other popular messenger, has many advantages. However, it is almost not protected from spam. Due to its prevalence, this combination attracted a lot of attention from spammers and created problems for communication. To add a person to spam mailing lists, it was enough to know their phone number, so millions of advertising messages were sent out every day. Users who are not on your contact list see a photo, IP address, and other personal information.

Don't forget that some of the company's data is stored in Russia. Although they claim that they only have information about Russian users, this information cannot be verified.

In 2022, Viber joined the EU Hate Speech Code to prevent and combat the spread of discrimination and illegal content. Viber monitors and removes content in public channels when users post comments with hate speech elements

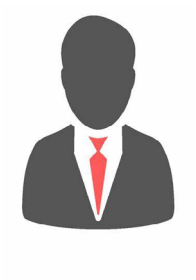
A *useful tool* – Viber Osint (GitHub, verification of Viber number registration Viber).

7. Formation of an Individual Profile

The starting point for the search can be any available information – a person's surname and name, mobile phone number or email address, social media profile, place of work / occupation, area of public activity, participation in certain events, photo, or video fragment, etc.

In other words, the formation of a **person's profile** (dossier) begins with a certain initial set of information that needs to be linked to other open data (if available) using search tools, dividing the information received into functional blocks in the reporting document:

Surname, first name and patronymic
(possible changes to full name, nickname, call sign)



Foto

Personal data (date and place of birth, place of registration (residence), passport of a citizen of Ukraine (foreign passport, citizenship of other countries), taxpayer registration number, entry number in the Unified State Demographic Register, driver's license), **contact information** (mobile phone, email address, Skype, messengers, IP address), **social media accounts**, etc.:

- ***search and meta-search engines***;
- ***Telegram bots*** for universal search;
- ***socially oriented platforms***;
- ***specialized services*** – Castrick (free – basic functionality), DarkGPT (GitHub, a tool based on ChatGPT-4 for searching for data leaks), Epieos (search by email or phone number, requires registration, free – Google, Email Checker & Skype, Clickable links, watermark), Hive (GitHub, automates data collection through Truecaller, Shodan, IntelX, Email Verifier, Sherlock, etc.), Phunter (GitHub, search by phone number), PrivacyWatch (free – basic functionality), Pipl (free – 5-day access; registration required), SpiderFoot (GitHub, search by IP address, domain, email or phone number), Spokeo (results are paid), Uscrapper Vanta (GitHub, collects and downloads email addresses, links to social networks, geolocation, phone numbers, user nicknames from the target site; can filter results by keywords), Webmii (information from social networks, websites and online documents), X-Ray (free – 2 credits, requires registration, search for Russians during the armed aggression is free);
- ***phone numbers*** – Moriarty Project (GitHub); mobile *applications for identifying subscriber numbers* CallApp, CheckerUA, Eyecon, Getcontact, NumBuster, TrueCaller, WhoCalls (by default, they download user phone book entries; for safe operation, a separate phone with simulated contacts is required);

	phone number databases – Spravkaru.net (RU), spravochnik109 (RU), Directory of city telephone numbers (RU, Russia, Ukraine, Belarus, Moldova, Latvia, Kazakhstan), Who Called?; *other* – IMEI.info (search by IMEI, International Mobile Equipment Identity), email2phonenumber (search by email), PhoneInfoga (GitHub, information about the phone number); • ***email address*** – GHunt (GitHub, collects various information about Google users; there is an online version that requires registration), H8Mail (GitHub, scans a specified mailbox and provides a list of possible passwords for it), PasswordSearchBot (bot, searches for emails and provides «merged» passwords, free – 10 requests per day), YaSeeker (GitHub, information about Yandex account by email or login), Zehef (GitHub, email research); *email validators* – Email Hippo, Verifalia (free – 25 checks per day), VerifyEmailAddress; email header analysis (sender's IP, mail servers, sending path) – Email Header Analysis, Messageheader, • ***State Migration Service of Ukraine services*** – Check against the database of invalid documents, Check for extension of stay/temporary; • ***Ministry of Internal Affairs of Ukraine services*** – search for a passport of a citizen of Ukraine among stolen and lost passports,checking an extract from the Unified Register of Persons Missing in Special Circumstances, Missing Citizens.
Marital status, family:	
	• ***state registers*** – Open Register of National Public Figures of Ukraine, Unified State Register of Declarations; • ***socially oriented platforms***; • ***dating sites*** (registration required) – Badoo, Jolly, Tinder, UkrDate; • ***family tree*** – FamilySearch (registration required); • ***obituaries*** – Legacy.
Education, academic degree, academic title, scientific publications:	
	• ***search and meta-search engines***; • ***specialized services*** – Google Academy, Science of Ukraine, Register of Documents on Higher Education, Ukrainian Science Citation Index.
Military, special, honorary titles, state awards:	
	• ***search and meta-search engines***; • ***specialized services*** – the website of the President of Ukraine (search by documents).
Professional activity, biographical data:	
	• ***state registers*** – State Register of Certified Forensic Experts, Unified State Register of Declarations, Unified Register of Advocates of Ukraine, Unified Register of Insolvency Receivers of Ukraine, Unified Register of Notaries of Ukraine, Unified Register of Private Enforcement Officers of Ukraine, Register of Certified Persons (architects, designers, experts, technical supervision engineers), Register of

Auditors and Audit Entities, Register of translators;

- ***participation in legal entities, opening a sole proprietorship***;
- ***job search sites*** – JOBS.ua, RABOTA.ua, WORK.ua and more;
- ***political and/or social activities*** – Database of politicians and parties, Financial statements of parties.

Property and financial status, income, vehicles:

- ***state registers*** – State Register of Real Property Rights, Unified State Register of Declarations;
- ***land plots*** – Cadastral map of Ukraine, CadastreService;
- ***vehicles*** – Autodetective, Baza-gai, Carma, Checkcar, Unda, AvtoNomera (search by state number and VIN code); Plate Recognizer (recognizes the brand, color, type of car and country of registration of the license plate; free – 2500 views per month, requires registration); Motor (Transport) Insurance Bureau of Ukraine (checks the validity of the insurance policy); Vehicles wanted; Information about vehicles and their owners;
- ***classifieds*** (including on local resources) – Domik (photo/video of premises/buildings), M2bomber (additionally - phone search), OLX.ua, Prom.ua, RIA.com, etc;
- ***intellectual property***;
- ***other*** – Binlist.net (identification of the bank by card number).

Court cases, compromising material, enforcement proceedings, conflicts:

- ***enforcement proceedings*** – Automated system of enforcement proceedings, Unified Register of Debtors;
- ***state registers*** – Unified State Register of Persons Subject to the Provisions of the Law of Ukraine «On Purification of Government», Unified State Register of Persons Who Committed Corruption Offenses, Catalog of Corruption Risks, System for Finding Hidden Interests (National Agency for the Prevention of Corruption, requires registration);
- ***activities to the detriment of the national security of Ukraine*** (according to NGOs) – Evocation.info (collaborators), Myrotvorets and IDentigraF (search by photo in Myrotvorets database, requires registration, 5 requests per day), Register of Traitors, ORDILO;
- ***pre-trial investigation*** – summonses, notices of suspicion and information on suspects in respect of whom permission to conduct a special pre-trial investigation has been granted on the website of the Prosecutor General's Office and the Uriadovyi Kurier newspaper;
- ***materials of journalistic investigations*** – Antikor, General Staff, social movement Chesno, Hroshi, Dossier, Our hroshi, ORD, Politrada, Slidstvo.Info, Slovo i Dilo, Schemes, Transparency International Ukraine (Anti-Corruption), Ukr.Av, Censor.Net, Anti-Corruption Action Center, Bihus.info, DocumentCloud (open database of political and legal documents used by journalists in their investigations),

	LittleSis (інформація про публічних осіб); • ***persons hiding from the authorities*** – Information on persons hiding from the authorities, wanted by the SSU and MIA; • ***sanctions*** – State Register of Sanctions, Consolidated Sanctions List of the UN Security Council, List of Persons Associated with Terrorist Activities or Subject to International Sanctions, Register of Individuals Under Sanctions of the National Security and Defense Council, Consolidated Canadian Autonomous Sanctions List, EU Sanctions Map, Office of Foreign Assets Control (OFAC, USA), OpenSanctions, SanctionsExplorer; • ***court decisions*** – Unified State Register of Court Decisions, Register of Court Decisions (requires registration), Status of Court Cases (search for a party to a case by name on the Judicial Power of Ukraine portal).
Immediate environment, friendships, hobbies, interests, habits, inclinations, regular and irregular places of visit, staying abroad, lifestyle, dominant needs, and other information that should be taken into account:	
	• ***search and meta-search engines***; • ***socially oriented platforms***; • ***court cases, dirt, enforcement proceedings, conflicts***.

8. Formation of a Legal Entity Profile

A legal entity is an organization that has undergone a *legally approved registration procedure*, a subject of law what has property rights and the ability to act as a plaintiff and defendant in court. *The mandatory attributes* of a legal entity are the following: a) constituent documents that reflect the system of governing bodies (sole, where there is one head; collegial, where decisions are made by several persons) that form and express the will of the legal entity, and units that perform certain functions and are established in the charter; b) authorized capital and a bank account; c) separate property, property and, in certain cases, subsidiary liability; d) the ability to act in court on its own behalf, indicating the organizational and legal form and individual name.

Each state has its own rules for legal entities.

Types of legal entities: 1) depending on the procedure of establishment – private law legal entities and public law legal entities; 2) depending on the main purpose of activity – commercial legal entities (carry out business activities for profit, and the profit is distributed among its participants), non-commercial legal entities (created to achieve social, charitable, cultural, educational, scientific and administrative goals, to protect the health of citizens, to develop physical culture and sports, to satisfy spiritual and other goals aimed at achieving public goods).

The **profile of the legal entity** should include its details (name, location, identification code, banking information), as well as:

Name of the legal entity

(full and abbreviated in Ukrainian, foreign language)

State registration (EDRPOU code, organizational and legal form, location, contact information, date of registration, founders/beneficiaries/authorized persons/employees, their changes, number of employees, amount and composition of authorized capital, corporate structure, branches, KVEDs, financial and tax reporting, bankruptcy/ termination status, potential fictitiousness, etc):

LOGO NAME
TAGLINE TEXT

Logo

- ***search and meta-search engines***;
- ***state registers*** – State Register of Printed Mass Media and Information Agencies, State Register of Scientific Institutions Supported by the State, Unified State Register of Legal Entities, Individual Entrepreneurs and Public Organizations, Unified Register of Public Organizations, Unified Register of Enterprises Subject to Bankruptcy Proceedings, Unified Register of Receivers and Receivers budget funds,

Register of public associations, Register of VAT payers;

- ***aggregator services of state registers*** – Clarity-Project, ContrAgent, E-data, Nomis, OdnodataUA, Opendatabot, VkursiPro, YouControl and more;
- ***податки*** – Registers of the State Tax Service of Ukraine (Data on registration of taxpayers, Register of insurers, Register of single tax payers, Register of taxpayers using a single account, Certificate of no debt, Data of the Unified Register of Individual Tax Consultations, Register of Non-Profit Institutions and Organizations, Data of the Register of VAT Payers, Search for Fiscal Receipt, Search for Excise Tax Stamp, Information on Cash Registers, Information on Cash Registers, Information on Cash Register Books, Cash Register Exemplars, Register of Cash Registers; some services require authorization);
- ***stock market*** – Stock Market Infrastructure Development Agency of Ukraine (SMIDA), State Register of Securities Issues, State Register of Authorized Rating Agencies, Register of Collective Investment Institutions, Register of Non-State Pension Funds, Register of Associations of Professional Capital Market Participants, Registers of law enforcement (issuers with signs of fictitiousness, absence at the location, prohibition of trading in securities on exchanges, etc.), Register of professional participants of capital markets and organized commodity markets, Register of certified persons, Securities of foreign issuers admitted to circulation in Ukraine;
- ***other*** – Consolidated List of Natural Monopolies, Importers and Exporters of Ukraine, Integrated Information System of the National Bank of Ukraine (State Register of Financial Institutions and Register of Persons that are not Financial Institutions but have the right to provide certain financial services online), LinkedIn (employees), Registers of the State Service of Ukraine for Food Safety and Consumer Protection.

Financial and economic activities (tenders, counterparties/related parties, export-import operations, intellectual property, business reputation, sanctions)**:**

- ***public procurement*** – .007, Anti-Corruption Monitor, Public control of state procurement, Unified Web Portal for the Use of Public Funds, State procurement, Summary of Information on Distortion of Tender Results, Register of the Antimonopoly Committee of Ukraine Decisions on Anti-Competitive Concerted Actions, Є-data (public finance), Prozorro, BI Prozorro, Public Bid, Zakupivli.pro;
- ***foreign economic activity/contractors*** – 52wmb.com (aggregator of customs invoices), Business registers EU (EU business registry), Dato Capital (private companies and their managers, free – basic functionality), Dun & Bradstreet (aggregator of legal entities registers, free – basic functionality), European data (open EU data), Eurostat (statistical information), EU tenders (public tenders of the EU), Global Tenders (public tenders more than 190 countries), ICIJ Offshore Leaks Database (offshore leaks), ImportGenius, ImportKey, ImportYeti (access to the US Customs Service database, requires registration), NBD Data, North Data (earch by European companies), OCCRP Aleph

(investigative journalism), OCCRP ID (a collection of resources for tracking companies and assets depending on the region, country or field of activity; paid access or registration is required), OpenCorporates (aggregator of registers of legal entities of the world, free – basic functionality), Open Ownership Register (beneficial owners), Opentender (public tenders 35 European countries), Trade Database Free (requires registration), UN Comtrade (global database of foreign trade statistics), Vat-search (VAT payers; free – 3 credits per month, basic functionality), Volza, Worldwide Registers (business registers of countries of the world), YouControl World (connections between companies and individuals from the CIS and the UK, free – 7 days); List of foreign trade registers (company registers);

- ***sanctions*** – sanctions against individuals, as well as the Database of legal entities subject to sanctions, Register of Sanctioned Companies of the National Security and Defense Council, Special Sanctions of the Ministry of Economy of Ukraine, List of companies that as of 24.02.2022 had an owner or beneficiary from Russia;
- ***property*** – State Register of Real Property Rights (request by EDRPOU code);
- ***intellectual property*** – State system of legal protection of intellectual property, Customs Register of Intellectual Property Rights, Special Information System of Ukrainian National Office of Intellectual Property and Innovation, Ukrainian National Office of Intellectual Property and Innovations, Iprop-ua.com, Opendatabot (trademark check, additionally – countries of the world); European Union Intellectual Property Office, World Intellectual Property Organization, U.S. Patent and Trademark Office (USA);
- ***court cases, compromising material, enforcement proceedings, conflicts***.

Availability of a license or special permit:

- ***energy*** – License Register of business entities engaged in economic activities in the field of heat supply, List of business entities that have obtained licenses for the production of heat energy at thermal power plants, thermal power plants, nuclear power plants, cogeneration plants and plants using non-traditional or renewable energy sources;
- ***forests*** – Registers of the State Agency of Forest Resources;
- ***medicine*** – License Register of the Ministry of Healthcare of Ukraine, Registers of the State Service of Ukraine on Medicines and Drugs Control;
- ***subsoil use*** – Register of concession agreements, special permits for subsoil use (scanned copies of special permits and use agreements);
- ***transport*** – License Register of International Transportation, License Register for the Conduct of Economic Activities for the Transportation of Passengers, Dangerous Goods and Hazardous Waste by Road, License Register for the Conduct of Economic Activities for the Transportation of Passengers, Dangerous Goods and Hazardous Waste by Rail, Register of Air Transportation Licenses;
- ***other*** – License Register of Economic Activities for the Provision of Firefighting Services and Works, List of Permits for the performance of high-risk works

	and for the operation (use) of high-risk machines, mechanisms, equipment, List of business entities that have licenses to carry out economic activities for the provision of services in the field of technical protection of information, Register of licenses for the use of radio frequency resources of Ukraine, Register of licenses for security activities and repair of firearms for non-military purposes.

9. Transport and Container Tracking

Aircraft: *trackers* – ADS-B Exchange, Adsb.fi, Flightradar24, FlightAware, IntelSky Military Radar, OpenSkyNetwork, Planefinder, RadarBox; *other* – Airframes.org (international aircraft registry), Aviation Safety Network (database of air accidents), Drone Crash Database (database of military unmanned aerial vehicles accidents according to media reports), FlightConnections (scheduled flights), Jetphotos.com (aviation photos), OpenSky-Network (flight tracking network), Planespotters.net (information about aircraft, their photos, flights, tracker, etc.), SkyVector (airborne software for flight planning); Orbitrack (real-time satellite tracker).

Seagoing vessels: *trackers*– MarineTraffic (free – 7 days), Marine Vessel Traffic, Military Ship Tracker (warships), MyShipTracking, SeaTracker (RU), ShippingExplorer, ShipTraffic, VesselFinder, VesselTracker, *other* – Balticshipping and Crewell (employment of seafarers), BoatInfoWorld (reference information on ships), Crew List Index Project (ihistorical information about ships and crews), LogisticsGlossary (glossary of logistics terms), Maritime Awareness Project (map of maritime borders and economic zones), OpenSeaMap (nautical map, ship traffic, weather and depth monitoring), Sea Ports Catalog (port catalogue), Shipspotting (reference information about the ships, their photos), World Shipping Register (ships and shipping companies data).

Container transportation: *trackers* – CMA CGM Group, Container-Tracking, Maersk Tracking, SeaRates Container Tracking, ShipIt Container Tracking, ShipmentLink, Shipping Container Info, ShippingLine, Track-Trace, Utopiax Container Tracking, Карта руху суден; *other* – BIC-Code (international register of container owners).

Land transport: geOps (online train tracker), OpenRailwayMap (map of railway infrastructure), Transit Visualisation (visualization of land transport), WikiRoutes (RU, public transport directory), Yandex.Timetables (online map of train, electric and buses in Russia, Belarus and Kazakhstan).

10. Tools to Counter Russian Aggression

Individuals and legal entities of the Russian Federation:

- catalog of services, portal and hub ***of the open data*** of the Russian Federation;
- ***a selection of resources*** for searching for data on individuals and legal entities (RU), OSINT Russia (a catalog of links from OsintFlow);
- ***citizens of the Russian Federation*** – Cybersec.org (and CyberSec Karma Bot), DataAnalytic (bot), Mail2Phone (RU, bot, can find a phone number by a known email address associated with Sberbank and Odnoklassniki profiles), OsintFlowFindBot (bot, free – 10 requests daily, basic functionality), Prob3y (RU, bot), Reveng.ee, X-Ray (free – 2 credits, requires registration, search for Russians during the armed aggression is free), Open database of data of public officials of Russia, Belarus and Kazakhstan, Information on individual tax number of an individual, Checking the validity of individual tax number of individuals (RU, the date of invalidation in most cases coincides with the date of death), Checking the validity of the passport of a citizen of the Russian Federation (RU), as well as bots for universal search;
- ***war criminals*** (information is not always official and therefore needs to be verified) – War and Sanctions, War Criminals of the Russian Federation, Book of Torturers of the Ukrainian People, "Collaborators and Traitors" (RU, YouTube channel), Myrotvorets and IDentigraF (search by photo in Myrotvorets database, requires registration, 5 requests per day), "Don't Wait for Me from Ukraine" (Telegram channel), "List of Corrupt and War-mongering People" (RU, A. Navalny's International Anti-Corruption Foundation), «Putin's List» (RU, database of the Free Russia Forum), Evocation.info (propagandists), Lostivan Wiki, Russian War Criminals, as well as *materials of OSINT investigators* (InformNapalm, Molfar, OsintFlow, OSINT Бджоли, Truth Hounds, etc.); *military equipment* – WarSpotting;
- ***vehicles*** – Nomerogram (RU), services of the traffic police of the Russian Federation (RU), AVinfoBot (RU, bot, requires a subscription), VIN01 (RU);
- ***financial and property status*** – Declarator (RU, non-governmental aggregator of property declarations), Information portal on real estate objects (RU), My taxes (RU), cadastral maps – Public cadastral map (RU), Debt Center (RU), Egrp365 (RU, unofficial analogue of the state register of real estate), ShtrafKZBot (RU, bot, checking fines/taxes/penalties);
- ***other*** – Database of deputies of the United Russia (RU), Service for checking invalid passports (RU), Dominfo.info (RU, search for information about developers, management companies and real estate);
- ***websites (registries) of public authorities*** – Prosecutor General's Office

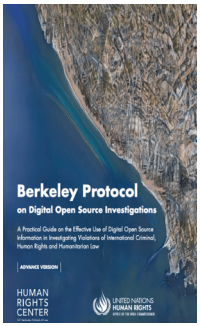
	of the Russian Federationї (RU); state register of companies (unified state register of companies) (RU), state register of real estate (RU), unified federal register of information on bankruptcy, debtors and auctions for the sale of collateral (RU), unified federal register of information on the facts of legal entities' activities (RU), cadastral map of Russia (RU), officially published legal acts (RU), register of unscrupulous suppliers (contractors, executors) and the register of unscrupulous contractors (RU), register of natural monopolies (RU), register of federal property of the Russian Federation (RU), federal tax service (RU, search by registers, the service "Transparent business", extract from the unified state register of legal entities/unified state register of enterprise), social insurance fund (RU); • ***checking counterparties*** – E-DOCIE (RU), ZaChestnyBiznes (RU, free – basic functionality), Updated information on joint-stock companies (RU), Chekko (RU), Audit-it (RU), Database for all Russian companies, DataNewton (RU), Egrul_bot (RU, bot), Fek (RU) (free – basic functionality), Injust.pro (RU), List-org (RU), Rusprofile (RU, free – basic functionality), Star-Pro (RU, free – basic functionality); • ***tenders*** – Unified Information System in the field of procurement (RU), RosTender (RU, free – basic functionality), TenderGURU (RU); • ***securities*** – Prime, SCRIN, Disclosure.ru (RU, disclosure of information on the securities market); registers of the Bank of Russia (RU); • ***intellectual property*** – open registers of the federal institute of industrial property (RU); • ***education*** – national accreditation agency in the field of education (RU), register of licenses (RU), register of organizations engaged in educational activities under educational programs that have state accreditation (RU); • ***courts, notaries*** – file of arbitration cases (RU) and cases of general courts (RU), portal of arbitration courts (RU); notary information portal (RU, enforcement proceedings, search for heirs), federal bailiff service (RU, enforcement proceedings).

Individuals and legal entities of the Republic of Belarus:	
•	***websites (registers) of public authorities*** – unified state register of legal entities and individual entrepreneurs (BY), unified state register of bankruptcy information (BY), unified register of licenses (BY), Ministry of Taxes and Fees (BY, registers of individuals and legal entities), Ministry of Transport and Communications (BY, carrier licenses); Ministry of Justice (BY, information on debts); national legal internet portal of the Republic of Belarus (BY), public cadastral map (BY), register of addresses (BY), register of real estate (BY), register of certificates of state registration of goods (BY), register of characteristics of real estate (BY), trade register (BY);
•	***counterparty verification*** – BizInspect (BY), Card Index (BY), StatusPro (BY);
•	***intellectual property*** – National Center for Intellectual Property (BY);
•	***courts*** – e-judicial portal (BY).

11. Using the Open Data for the Purposes of Pre-trial Investigation. Berkley Protocol

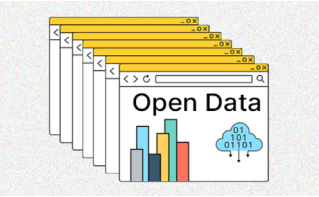
The digital technology era consistently sets new requirements for law enforcement agencies to effectively fulfill their tasks during pre-trial investigations. Under these circumstances, IT processes necessitate the adoption of new approaches for the proper collection and preservation of digital information from open sources, which can subsequently be used as evidence in criminal proceedings.

One of the greatest challenges that they face is dealing with the discovery and verification of relevant material within an increasing volume of online information. In order to implement the experience of the best international practices, while strictly complying with the requirements of the Criminal Procedure Code of Ukraine (hereinafter – the CPC of Ukraine), the Office of the Prosecutor General recommends in 2021 that the principles, methodologies and standards for reviewing digital information from open sources set forth in the Berkeley Protocol be introduced into the practice of investigators and their procedural supervisors.



The Berkeley Protocol (informally translated into Ukrainian) is a practical guide on methods and procedures for using publicly available digital information in investigating violations of international criminal law, human rights and humanitarian law, which was presented in 2020 by the UC Berkeley Human Rights Center (California, USA) and the Office of the United Nations High Commissioner for Human Rights. More than 150 international experts worked on it. It outlines the minimum standards for searching, collecting, storing, verifying and analyzing data from open sources in compliance with professional, legal, and ethical principles.

According to the conclusions of the Office of the Prosecutor General, this practice can be used not only during the pre-trial investigation and criminal proceedings on crimes of an international nature committed in conditions of extraterritoriality or those that are investigated without access to the territory of their place of commission, but also for collecting information from open sources that can be of evidentiary value in any category of proceedings.



For these purposes of the Berkeley Protocol (§ 14-18) **information in open access** includes publicly available data that anyone can *observe* (via a website using any free web browser), *purchase* (paid services that are available to all members of the public, not just certain groups, such as law enforcement officers or private private detectives) or *request* (appeals that may be submitted by any person regarding public information to State entities in accordance with regulations on freedom of information circulation or access to it),

without requiring special legal status or unauthorized access. It does not refer to requests to individuals, companies or organizations to voluntarily hand over their information, but is limited to requests to State entities.

Today, there is a growing amount of data on the Internet that is made public without the consent of its owners, either through hacking, leakage, security vulnerabilities, or publication by others without authorization. Although this information is publicly available and thus formally considered open, there may still be legal and ethical restrictions on how it can be used. In addition, digital information may be accessible to those with specialized technical knowledge and the ability to connect to networks and data that are not available to the average person (e.g., the Dark Web can only be accessed with certain programs, such as the Tor browser).

The Berkeley Protocol includes this information within the realm of open access, provided there is no unauthorized access to it. A key feature is that open access information does not involve interaction or requests for data from individual internet users. Obtaining information from individuals through direct communication is classified as a closed source. Thus, there is a general prohibition of unauthorized access to data and networks (e.g., using passwords, deception, or other forms of social engineering) (§ 63 of the Berkeley Protocol).

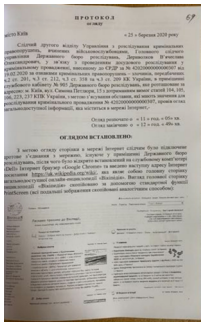
Closed-source information refers to the *data with restricted access* or *protected by law*, which can only be obtained legally through private channels, such as court proceedings, or voluntarily provided by an individual. Acquiring information from other Internet users through communication with those users is considered closed source.

According to paragraph 65 of the Berkeley Protocol, the use of virtual personas violates the terms of the user agreement of services and, in particular, social media platforms. Although such virtual identities are necessary when used to search for and store data in the public domain, they should not be used to attempt to collect content posted on social media that is restricted for free viewing; or as a pretext to obtain information directly from a person under the guise of a false identity. Such behavior, according to international experts, takes researchers beyond the scope of open data investigations, contradicts ethical principles, and may violate relevant legal provisions (the right to privacy and data protection, etc.).

The term «evidence» should be distinguished from «information». Open source evidence is open source information with evidentiary value that may be admitted in order to establish facts in legal proceedings. Once digital content has been identified and deemed relevant for investigation, the investigator must determine the appropriate method of collection. This method may vary depending

on whether the online content has potential evidentiary value in judicial proceedings, will be used for decision-making, or will contribute solely to internal outcomes. In cases where the work focuses solely on internal results, a screenshot or converting a webpage to a PDF file may suffice. However, content with potential evidentiary value may require a more thorough and justified method of collection (§ 153 of the Berkeley Protocol).

The Law of Ukraine "On Amendments to the Criminal Procedure Code of Ukraine and the Law of Ukraine "On Electronic Communications" to Increase the Effectiveness of Pre-trial Investigations "in Hot Pursuit" and Counter Cyberattacks" dated March 15, 2022, № 2137-IX Article 237 CPC of Ukraine introduced *a new object of inspection* – computer data (part 1) and *specified the requirements for its examination*. The inspection of computer data is conducted by the investigator or prosecutor by recording in the inspection protocol the information contained therein in a form suitable for perception (using electronic tools, photography, video recording, screen recording, and / or video recording of the screen, or in paper form) (§ 2 of Part 2).


ПРОТОКОЛ

ОГЛЯДОМ ВСТАНОВЛЕНО:

Simultaneously, courts, aiming to adhere to the principle of direct examination of evidence obtained through website inspection, raise questions about studying the online resource from which the respective information copy was made during the court hearing. However, this may not always be feasible due to the possibility of deletion or modification. Therefore, to ensure the accessibility of information from open sources, digital storage (archiving) should be carried out. Archiving allows for the preservation and protection of data over time, including its authenticity, accessibility, identity, permanence, rendering (visualization), and clarity. These indicators of digital information are subject to documentation per the Berkeley Protocol.

When performing the aforementioned investigative actions in accordance with Articles 223 and 237 of the CPC of Ukraine, it is mandatory, under Article 71 of the Code, to *involve a specialist* with higher education in the field of information systems and technologies. This is required for the subsequent court recognition of the information contained in the created electronic document as an original according to Part 4 of Article 99 of the CPC.

Investigators conducting investigations using open-access data must collect online content in its native format or as close to its original state as possible. Any changes, transformations, or conversions caused by the collection process must be documented (§ 154 of the Berkeley Protocol).

While collecting all the below-mentioned information is considered the best practice, the **first three points** are the **minimum standard for presenting evidence**

in court (paragraph 155 of the Berkeley Protocol):

(a) *Target web address: the web address of the collected content, also known as the uniform resource locator (URL) or identifier (URI).*

URL (Uniform Resource Locator) is the address of a resource on the Internet. It includes the protocol name (https, ftp, telnet, gopher, etc.) and the path to the resource, formatted according to the access scheme. For example, https://www.president.gov.ua/ (the official website of the President of Ukraine).

The URL standard uses US-ASCII characters, i.e., Latin letters, digits, and certain punctuation marks. All other characters (including Cyrillic letters) must be encoded per the URI (Uniform Resource Identifier) protocol. For example, a string like https://uk.wikipedia.org/wiki/Вікіпедія displayed by the browser as https://uk.wikipedia.org/wiki/%D0%92%D1%96%D0%BA%D1%96%D0%BF%D0%B5%D0%B4%D1%96%D1%8F (so-called percent encoding).

After describing the sequence actions of the investigator's to find the content relevant to the investigation on the Internet, the target web address is recorded by *copying the original URL* from the browser address bar (if the transition to it was redirected, then both the start and end addresses) and including it in the text of the procedural action protocol. We do not recommend using link shortening services (Bitly, is.gd, Short URL, Surli, Tinyurl та ін.) due to the unclear algorithms of their operation, the lack of data on the duration of such links and guarantees of their immutability.

In order to make it easy to follow the URL and view the content of the target web page, it is a good practice to display, in addition to its textual address, the following *QR code* (quick response code). To do this, use the built-in functionality of the browser (e.g., in Chrome, right-click on an empty spot on the page, select «Create QR Code for this page» and copy it) or specialized resources such as MeQR, Online QR Code Generator, QRcode Generator, QRCode Monkey. The required data are then extracted from patterns that are present in both the horizontal and the vertical components of the QR image.

(b) *Source code: investigators must capture the HTML source code of the web page, if applicable. HTML source code includes a lot more information than the visible portion of the website. The HTML source code will contribute to the authentication of the material collected.*

HTML (HyperText Markup Language) is a standardized markup language used to display web resources in a browser during their loading. With HTML, you can publish not only text on website pages, but also multimedia content (video, photos, audio), mathematical formulas, and other objects.

HTML code is processed by the browser as text documents with the extension .htm or .html. Typically, this is a list of numbered lines with information about a particular site element.

To *download the HTML code of a webpage*, right-click to open the context menu and click «Save as» (or use the keyboard shortcut Ctrl+S for Windows or Cmd+S for Mac OS). After this, you need to choose one of the possible saving options: *only HTML* (only the code without additional content; the page may later display incorrectly), *one file* (in MHTML format; optimal solution in most cases), or the *full webpage* (in addition to the HTML code, a separate folder is saved with all its elements, including photo and video files; at the same time, it is necessary to describe all files in the protocol) and briefly display this action in the procedural document with the name(s) of the file(s).

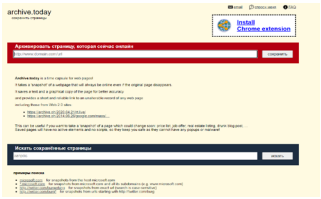
Given the technical features of individual webpages, the Berkeley Protocol allows not to download their HTML code if it is impossible. In such cases, this must be noted in the investigative action protocol.

(c) *Full-page capture: investigators should first take a screen capture of the target web page with the date and time indicated. The reason for this process is to have the best possible representation of what was seen at the time of collection.*

A screenshot can be taken using *built-in OS tools* – the PrtSc key, Print Screen, Snipping Tool (for Windows), or Shift+Cmd+4 (for Mac) or *using programs* like Apowersoft Free Screen Capture, FastStone Capture, Greenshot, LightShot, PicPick, ShareX. Screenshots must reflect the entire sequence of actions of the investigator (especially for nested pages), contain the system time and date corresponding to the time and date of the review. For security reasons, it is advisable to remove screen elements unrelated to the subject of the review (folders, program icons, other open tabs, personal information, etc.) in advance, or create a separate virtual desktop.

If the *target webpage contains scrolling* (new parts of the content dynamically display when the user scrolls the page, such as in social media or messengers), you can: 1) take several screenshots with partial overlap; 2) take a so-called long screenshot using Apowersoft Free Screen Capture, FastStone Capture, PicPick, ShareX or the «Developer Tools» feature in the Chrome browser (Ctrl+Shift+I → Ctrl+Shift+P «Run Command» → start typing «screen...», a suggestion for «Capture full size screenshot» will appear, select this command and wait for the file to be saved in .png format); 3) export the page in .pdf; 4) record a video of its viewing using screen recording programs. In this case, a prerequisite is to first load the necessary part of the resource (i.e., the actual loading of the dynamic site section to the location, for example, of the target post).

In addition to the three mandatory points outlined in the Berkeley Protocol, the Office of the Prosecutor General of Ukraine recommends **creating an archive of the target webpage** using internet resources designed for web archiving. These



services are electronic libraries that ensure the long-term preservation of collected material and constant access to them. Given the lack of control over the servers of such web archives, and therefore the inability to guarantee the immutability of data and unrestricted access, the best practice is to archive using multiple is to archive using multiple resources, such as Internet Archive (The Wayback Machine) and Archive.today (or one of its mirrors, archive.is, archive.li, archive.ph, archive.fo). In this case, the time and date of the archiving should match the time and date of the investigative action, i.e. the state of the target webpage is preserved exactly at the moment of the review. Using screenshots or a brief description in the text of the protocol, the archiving actions are documented, and a link/QR code to the created web archive is provided (best practice).

(d) *Embedded media files: if downloading a web page with videos or images, for example, those specific items should also be extracted and collected from the web page.*

Downloading embedded media files can be done using: 1) full saving of the webpage (see point b); 2) right-clicking on the image in the browser context menu and selecting «Save image as»; 3) using specialized services – FetchV (a Chrome plugin that downloads videos from websites where this function is not enabled by default), HImage (downloads all images from the page), as well as resources for video searching and verification and working on socially-oriented platforms.

Products for automated *creation of video transcripts* include Buzz (GitHub, supports Ukrainian), Happy Scribe (free – 30 minutes with registration on the website), Sonix (free – 30 minutes with registration on the website), Trint (free – 3 3 files up to 3 hours, valid for 7 days) Whisper (GitHub), Whisper WebGPU.

(e) *Embedded metadata: investigators should collect the additional metadata of the digital item, if available and applicable. Metadata can vary depending on the sources, but common metadata include uploader user identifier; post, picture or video identifier; uploaded date and time; geotag; hashtag; comments; and annotation.*

Embedded metadata is important for describing the digital content, the circumstances of its creation, distribution, or modification (for example, for an office document, this might include the date and time of creation or last modification or copying to a specific storage device; the name of the user who created the file or made the last changes; file size, etc.). They can either be part of the file, displayed on the webpage, or included in its source code. Any parameter from the listed and unlisted ones has a specific category and format. The key factor in metadata is a clearly organized structure that allows both humans and technology to read the data. Such a distinction enables you to work with a huge amount

of information in a short period of time, use the received metadata for collecting, storing, searching, processing and combining in an automatic mode.

To *form a tabular description* of a large number of downloaded files and automatically retrieve embedded metadata from them, the best practice is to use specialized software – Directory Lister (free – 30 days; allows getting hash sums for files and folders), Filelist Creator, MediaInfo and others.

(f) *Contextual data: contextual content should also be collected if it is relevant to understanding the digital item. This may include comments on a video, image or post; upload information; and/or uploader/user information, such as a username, real name or biography. Whether surrounding information should be collected needs to be determined based on the specifics of the case and the digital item.*

Contextual data can also include information about user is interactions with the posted materials – the number of views, citations, reposts, likes / dislikes, etc.

(g) *Data collection: open source investigators must record all relevant data pertaining to the collection, such as the name of the collector, the IP address of the machine used to collect the information, the virtual identity used, if any, and a time stamp. Investigators should make sure that the system clock is accurate, preferably by synchronizing it with a Network Time Protocol server. The reason for this step is to ensure that time-related metadata are accurately represented in the collected files. If a virtual identity is used to access the collected information, that should be noted.*

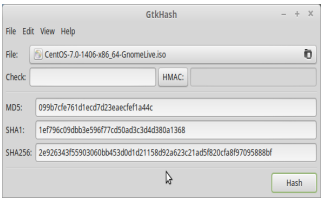
In the introductory part of the investigative report, it is advisable to indicate the date and time of the beginning and end of the inspection (suspension and resumption), information about the person conducting it (if necessary, a reference is made to the execution of the investigator's order); number, date and qualification of the criminal proceedings; place of execution; identifiers of the computer equipment involved (laptop, printer, optical drive, etc.); the operating system installed on the personal computer, the name and version of the browser, as well as other necessary programs (in particular, VPN); participants in the investigative action, the environment of its execution, etc.

(h) *Hash value: hash value is a unique form of digital identification that confirm, through the use of cryptography that the content collected is unique and has not been modified since the time of collection. At the point of collection, open source investigators should manually add – or the collection tool should automatically add – a hash value. There are many different types of hashes to choose from and the standards have evolved over time. Investigators should evaluate which hash to use based on the currently accepted standard.*

A hash sum (or hash value) is a fixed-length sequence of characters obtained by transforming arbitrary input data (numbers, text, files, etc.) using a special

mathematical algorithm. Hash functions are used to verify their integrity during transmission or storage (i.e., protection against modification). The process of transforming data into a hash is called hashing, and the hashing algorithm is called a hash function. Most common hash functions output large numbers in hexadecimal representation (for example, the SHA-1 algorithm results in 7DD987F84 6400079F4B03C058365A4869047B4A0). The hash value can be used to verify data integrity, identify and search for data (for example, in p2p networks), and replace the data that is not safe to store explicitly (passwords, answers to test questions, etc.).

Hash sum properties: *irreversibility* (the original data cannot be retrieved from it by mathematical methods or brute force), *reproducibility* (transforming the same input data using the same hash function always results in the same output), *uniqueness* (when hashing different input data, different hashes are generated, even if the data differs by only one bit).



Popular hashing algorithms include MD5 (hash length – 128 bits, recognized as insecure in 2011 due to a high likelihood of collisions, but still used for checking content integrity); SHA-1 (hash length – 160 bits); SHA-2 (hash length – 224, 256, 384, and 512 bits, with SHA-256 used in blockchain technology for transaction verification; works 2-3 times slower than MD5 and SHA-1) *and tools* – HashTab (PC software, no updates since 2022), GtkHash (GitHub), Hash Calculator Online (file size limit – 32 MB), Hash Checker (PC software), Hash Generator (PC software), Hash Tool (PC software), RapidCRC, (PC software, no updates since 2005), RHash (GitHub). The situation where two different data sets result in the same hash is called a collision. To properly distinguish such files, the best practice is to additionally specify the type and size in bytes for each of them. The obtained data should be entered into the protocol (in tabular form).

Therefore, the hash value must guarantee that the file was identified during the review, noted in the investigative protocol and its appendices, it has not been altered since collection, and can thus be considered entirely original and, therefore, admissible and valid evidence in the case.

Rules for assessing the admissibility of electronic evidence are outlined in the rulings of the Joint Chamber of the Criminal Cassation Court within the Supreme Court of Ukraine, dated March 29, 2021, in case No. 554/5090/16-k and September 25, 2023, in case No. 208/2160/18.

It is unfounded to equate electronic evidence as a means of proof with the physical carrier of such a document. A characteristic feature of an electronic document is the absence of a strict connection to a specific physical carrier. If it is stored on multiple electronic storage devices, each electronic copy is considered an original electronic document. The same electronic document can exist on

different carriers. All identical copies of the electronic document can be considered originals and differ from each other only in the time and date of creation. The issue of identifying the electronic document as an original can be resolved either by the authorized person who created it (by considering the file or directory checksum (CRC sums, hash sums) or applying a digital signature) or, if necessary, through conducting special judicial investigations.

For the purposes of criminal proceedings, the admissibility of an electronic document as evidence cannot be contested solely on the basis that it is in electronic form (part 2, article 8 of the Law of Ukraine «On Electronic Documents and Electronic Document Management»).

According to paragraph 157 of the Berkeley Protocol, the durability and accessibility of information on the Internet is often a subject to unforeseen circumstances - it can be easily decontextualized, lost, erased or damaged. The task of preserving digital materials is to ensure their integrity and unimpeded access. However, when it comes to preserving open data for legal accountability, the goal is to manage and preserve digital materials with a view to ensuring their accessibility, authenticity and usability in the accountability process, including their admissibility in court proceedings. Thus, preserving open data in the context of an investigation involves preserving information for a long time in such a way that the collected materials remain understandable to potential users regardless of the context and has a sufficient level of authentication.

In this regard, *all files preserved during the review must be recorded on a physical storage medium* (it is considered good practice to use an optical disc, such as CD-R, with a serial number), which, according to Article 105 of the Criminal Procedure Code of Ukraine, is attached to the protocol of the investigative action as an integral part.

To ensure that digital materials remain accessible and usable for the purposes of pre-trial investigation, it is advisable to follow the recommendations outlined in this section during their fixation. This should be carried out within the framework of counterintelligence and/or operational investigative activities through the preparation of *a review act*.

12. The Basics of Cryptocurrency Transactions Research

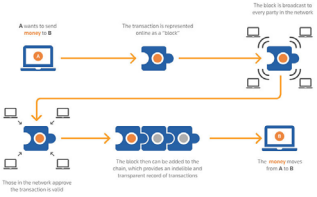
Cryptocurrency is a digital representation of value. This value can be the subject of digital trading and function as a medium of exchange, like the regular money we carry in our wallets. This type of property can be transferred to others or stored and traded electronically. Cryptocurrencies allow people to buy and sell goods without using the banking system, as they are not issued by central banks.

These operations are possible on a global level. They make national borders or national currency of the seller and buyer completely irrelevant. Typically, if there are transaction fees, they are quite low, and operations are executed very quickly, as no legal rules, formalities, or restrictions are applied. Finally, these values are fully decentralized (i.e., not tied to any national currency system), ensuring a certain level of anonymity.

The first successful *coin* (монетою) that gained widespread use as a means for payments, transfers, exchange, and accumulation was Bitcoin (BTC), created in 2009 by an anonymous user under the name Satoshi Nakamoto. All other coins that appeared after it (according to CoinMarketCap currently there are about 10,000), are called *altcoins* (alternative coins), such as Ether (ETH), Ripple (XRP), Cardano (ADA), Solana (SOL), Polkadot (DOT), Litecoin (LTC), Tron (TRN) and more.

The most important feature of a coin is the presence of its own *blockchain* – a decentralized digital ledger or special database that is supported by numerous computers located around the world (nodes). Blockchain data is organized into blocks, which are arranged in chronological order and protected by cryptography. Each block contains a timestamp, a hash (checksum) of the previous block, and transaction data presented as a hash tree. Transaction information is usually provided openly, unencrypted. Protection against forgery and tampering is ensured by including the hash of the entire block in the next block. Therefore, altering one block requires corresponding changes in all subsequent blocks, which is virtually impossible to achieve.

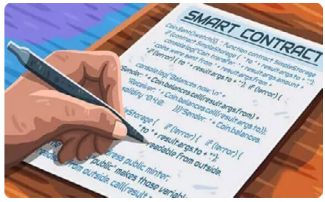
Such a distributed database is the basis of the first cryptocurrency, bitcoin, for which blockchain technology was created in 2008. In fact, it is a kind of ledger of all transactions that allows solving the issue of double spending without a central server or authority.



To sell cryptocurrency, for example, the seller initiates a transaction. It is transmitted to the blockchain network, where each node (computer) receives information about the transaction. Then, the nodes begin the process of verifying its authenticity

using consensus algorithms. After the nodes approve the transaction, it is added to a new block along with other recently approved transactions. The completed block is then added to the existing blockchain in chronological order. Each block contains a unique hash of the previous block, creating an unbroken and immutable chain. After adding a new block, all copies of the blockchain on the network nodes are updated to reflect the latest changes. Once a block is added to the chain, all its transactions are considered confirmed and irreversible. Verification of all transactions, including previous ones, takes place in every consensus cycle of the network.

Blockchain addresses where cryptocurrencies are stored and built on two keys – *public* and *private*. The public key is used for the «open» part of the address, while the private key is used to sign transactions and access the address, and it is intended only for the owner. The current computational power does not allow for breaking the blockchain address, specifically «guessing» the private key through brute force. Blockchain addresses and their management can be created and controlled in a special application – *a cryptowallet*. Wallets are based on various technologies, including blockchain, public key cryptography, and encryption.



The next stage in the development of virtual assets was the launch of smart contracts based on the Ethereum blockchain in 2015. A *smart contract* is a computer algorithm designed to enforce self-executing agreements, which will be ensured by the blockchain. Thanks to this, it became possible to issue an unlimited number of crypto assets and program their functions. This led to the creation of *tokens*. Smart contracts contain the balance values on the owners' accounts, which enables transferring them from one account to another without external intermediaries such as banks or government bodies. Furthermore, such transactions are traceable, transparent, and irreversible. A token does not have its own blockchain, which is its main difference from a coin. Smart contracts not only contain information about the obligations of the parties and penalties for their violation but also automatically ensure the execution of all contract conditions. In this interpretation, a smart contract is not necessarily related to the classical concept of a contract, but can be any computer program.

In essence, a *token* is a digital certificate, similar to securities (shares) used in the world of fiat currencies. It records the issuer's obligation to the token holder. Additionally, it is a unit of calculation that functions on other platforms. The scope of token use is broader than that of coins: they are used to provide services; to certify and confirm rights to something within the online platform itself; they

can be exchanged for other services or sold for another currency; and they are a powerful investment tool for startups, allowing their owners to receive dividends.

Thanks to the development and simplification of smart contracts, the most popular blockchain platform for tokens is Ethereum. Tokens based on this blockchain have a standard set of agreed-upon functioning rules ERC-20. The most popular ones, such as Shiba Inu, Tether, Uniswap and ApeCoin were created according to the ERC-20 standard. Other widespread token standards include BEP-20 (blockchain Binance Smart Chain або BSC, which has the same architecture as ERC-20), TRC-20 (blockchain TRON), ERC-721 (allows creating non-fungible tokens (NFTs) in the Ethereum network), ERC-777 (ca standard for fungible tokens, improving ERC-20), ERC-1155 (allows grouping transactions).

Although these cryptocurrencies offer a range of advantages, primarily the absence of the need to trust a bank for making payments anywhere to anyone, one of the key drawbacks is that their prices are unpredictable and tend to fluctuate, often significantly. This makes it difficult for ordinary people to use them. Typically, people expect to be able to control how much their money will be worth in a week, both for security and to ensure their means of livelihood. One of the ways to stabilize the exchange rate has been to tie coins to real assets (such as the US dollar, securities), exchange-traded goods (gold, oil), or other cryptocurrencies, with the creation of a centralized reserve for their guaranteed exchange at the rate (the difference fluctuates within 1%). This gave rise to *stablecoins* – Tether (USDT), USD Coin (USDC), TrueUSD (TUSD), Binance USD (BUSD), DAI, Tether Gold (XAUT), FRAX. Stablecoins have an obvious disadvantage: the level of trust in a centralized exchange mechanism required to maintain the exchange rate is at odds with the decentralized nature of cryptocurrencies.

NFT (Non-Fungible Tokens) used to confirm ownership and prove the authenticity of a particular virtual asset. If one Bitcoin is always equal to another Bitcoin, this is not the case with NFTs. Each such token is unique, with its own price and value. Experiments with NFTs began in 2013-2014 with the release of artworks, music, objects for blockchain-based games, collectibles, and more. As of September 2023, more than 95% of NFTs had a zero monetary value.

A cryptocurrency wallet is software that allows cryptocurrency transactions. It is important to understand that coins are not stored directly in your browser or on your computer; they are only displayed while being part of the blockchain network. To access cryptocurrency in a wallet and perform transactions

with it, two encryption keys are necessary – public and private. Without these keys, it is impossible to open access to the blockchain ledger and transfer the asset to another person.

A public key or *open key* is the address (account) where digital currencies are sent. It can be compared to a bank card number. The public key is used to create a transaction and can be shared openly if you expect funds to be transferred to your account.

A private key or closed key is the tool used to confirm transfers, similar to a digital signature. Without this key, the owner of the crypto wallet cannot manage its contents. The private key is often compared to the password for a bank card and must not be disclosed.

Each of the keys is a unique set of characters that includes letters and numbers. An example is the encryption keys of the Bitcoin blockchain. It is based on the SHA-256 algorithm that generates a 256-bit number. For more convenient operation, it creates a combination consisting of 64 characters – for example, 4BBFF74C A25A2A00409DCB24EC0418E9A41F9B3B56216A183E0E9731F4589DC6. This is the private key, although the length makes them extremely inconvenient to store, protect, and use. This problem becomes even more significant when interacting with multiple accounts, which involves recording and secure offline storage of several such combinations, as each cryptocurrency account inside your wallet requires a separate private key to manage. The latter allows the user to sign transactions, thus confirming agreement with the parameters of each transfer.

However, wallets have a *seed phrase or recovery phrase*, which is a unique mnemonic combination of 12, 18, or 24 words that acts as a backup for the crypto wallet. More precisely, the recovery phrase represents a long random number or entropy. Although the base of this number is random, the seed always includes words from a list of 2048 possible words in English (or the BIP39 list).

In essence, the seed phrase is the master key to all your private keys. Unlike private keys, the seed phrase does not allow signing transactions, but it provides instant access to every private key inside your crypto wallet and, therefore, to every account in it. If access to the wallet is lost, coins and tokens remain within the blockchain, while entering the seed phrase into another wallet in the correct order restores all private keys stored in the original crypto wallet. However, if this phrase is lost, it becomes impossible to manage the crypto assets.

Can I change an existing seed phrase? No, but you can create a new wallet and therefore a newseed phrase, and move funds to this new wallet. If your seed phrase is compromised, it is important to move the funds to a new address under the control of the new seed phrase as soon as possible

Crypto wallets are divided into custodial and non-custodial wallets. Non-custodial wallets can be further categorized into "hot" (desktop, mobile, online)

and «cold» (hardware and paper) wallets.

Creating a *custodial wallet* is similar to opening a bank account: the owner's data is transferred to and stored by a third party – the custodian (e.g., centralized crypto exchanges, some exchangers, and custodial services). To create one, you need to choose a company and register on its website, requiring an email or phone number and a password. Custodians have full control over the crypto assets and can intervene at any time (e.g., freeze accounts), but they are also responsible for safeguarding the client's funds.

A *non-custodial wallet* is an interface for accessing a wallet on the blockchain. The characteristic of *"hot"* storage is that such a wallet is always connected to the internet, and the key is stored in a desktop program or mobile app e.g., AtomicDEX, Blockchain Wallet, Coinbase Wallet, Exodus, MetaMask, Phantom, TrustWallet or in an online service's personal account – BitAddress, WalletGenerator. This allows for quick transactions but carries the risk of loss or theft, which could grant unauthorized individuals access to your cryptocurrency. A *"cold"* wallet is a more secure method of storing assets, as it allows keys to be stored outside applications or websites. It can take the form of a hardware device e.g., CoolWallet, Ledger, SafePal, Trezor, Walletz (resembling a USB drive or a bank card with Bluetooth, NFC, or QR code support) or a paper medium with the seed phrase or QR code.

A single crypto wallet allows working with multiple accounts. As the popularity of cryptocurrencies continues to grow, the number of transactions on blockchain networks can lead to scalability issues such as slow transaction times and high fees. This can affect the usability of cryptocurrency wallets and the overall user experience.

Various **ways to buy/sell cryptocurrency** include:

• *crypto exchange* an online service that based on blockchain technology, enables clients to buy, sell or store cryptocurrencies for other assets, fiat money, or other digital currencies. Users can create accounts, trade using various tools, open deposits, and withdraw cryptocurrency to external wallets or or bank accounts.

Centralized exchanges (CEX) often offer more services but a more modest list of available cryptocurrencies. They operate legally and guarantee the security of users' assets. However, such platforms will require mandatory verification. The exchange acts as a trusted intermediary between buyers and sellers. These platforms, like Binance, Bitfinex, Bitget, Bybit, Coinbase, Gate.IO, HTX, Kraken, KuCoin, OKX, WhiteBIT (UA).

Decentralized exchanges (DEX) operate in the "gray zone" and are not subject to the jurisdiction of a particular state. Unlike traditional CEXs, such platforms automate transactions and trading through smart contracts and decentralized

applications – users trade directly with each other, without the mediation of a centralized platform. They do not store keys to users' wallets, so they are not responsible for the safety of assets. The most famous DEX cryptocurrency exchanges – ApeX Pro, Balancer, Curve, dYdX, KyberSwap, OKX DEX, Slingshot, Uniswap, 1inch;

• *cryptocurrency exchanger* is a service that offers services for the exchange of digital assets, including fiat (state) money and vice versa. There are *three types of exchanges*: *online platforms*, *physical locations*, and *P2P exchanges*. The principle of operation is generally the same: the client submits an application, gives one cryptocurrency and receives another or money (or vice versa). The difference lies in the method of receiving the fiat: an online exchanger, for example BestChange (RU), Bitcoinmarket.global (RU), Bitcoin24, bits.media (RU), Changeit, ChangeNOW, Coin24, ObmenAT24, Obmify, Scanbit, 100btc, CourseExpert (RU), transfers it to a bank card and the physical one gives out cash. P2P services Bitcoin Global, Binance P2P, ByBit P2P etc. act as an intermediary between two clients, one of which buys cryptocurrency and the other sells it.

Unlike cryptocurrency exchanges, exchangers do not store the assets of users. They receive currency from the client to their wallet, and instead fiat (hryvnias, dollars, euros, etc.) from their own reserves. At the same time, the cryptocurrency exchange rate on such services already includes a commission for their services, which is why its price differs from the market rates;

• *crypto ATMs* are not linked to bank accounts. Instead, they are directly connected to cryptocurrency exchanges via the blockchain to enable instant buying and selling. These exchanges also determine the exchange rate based on the current market value at the time of the transaction. You just need to scan the QR code, transfer the cryptocurrency to the address provided, and then receive the money in cash or by bank transfer. For the latter, you need to insert your debit or credit card into the crypto ATM. The main disadvantage of crypto ATMs if you use them to withdraw cryptocurrency to a card is a significant fee. ATMs can be divided into *two main types*: *one-way* and *two-way*. The first type of ATM allows you to buy cryptocurrency only, while the second type offers an additional opportunity to sell it.

They can be located using services like Bitcoin ATM Map or Coin ATM Radar.

Transaction tracking services are platforms that have various tools for displaying cryptocurrency payments or withdrawals. Blockchain analysis involves research, classification, and monitoring of addresses and transactions.

However, it would be incorrect to say that the use of cryptocurrencies ensures complete anonymity, since the open blockchain technology used stores records of all transactions, and therefore they are available to other users, although this information does not contain personal data of their participants. For this reason,

cryptocurrency transactions are becoming increasingly popular for illegal activities.

Virtual asset transactions are a fairly broad source of information. Properly tracking them can help identify a suspicious person or transaction: OSINT technologies that use information from the blockchain in combination with appropriate software that can link crypto addresses to a specific centralized crypto exchange, exchanger, and even a specific user.

A person and a wallet are different concepts: "person" is not necessarily one person, it can be an organization, and "wallet" means that one person can have several wallets and several people can have access to one wallet.

Tools for tracking transactions – Arkham Intelligence, Blockchain Explorer, Blockpath, Breadcrumbs (free – 2 queries and 1 alert; registration required), Crystal Lite, GraphSense (GitHub), MetaSleuth (free – 200 queries per month), Orbit (GitHub), Shard (RU), Tokenview, WalletExplorer, Wallet-Tracker (GitHub); *for ETH* – Etherscan, Ethtective; *for TRX* – Tronscan; *for BNB* – Bscscan; *other* – Bitinfocharts (cryptocurrency statistics), Crypto Sanctions Screening Tools (sanctions in the crypto ecosystem). A drawback of some of these programs is the inability to visualize the transaction chain of a crypto wallet. More advanced tracking products – Chainalysis, Crystal Expert, Elliptic, Global Ledger, TRM etc. – require paid subscriptions.

Further identification of crypto wallet owners is usually carried out as part of procedural activities through submitting requests to specific exchanges or exchangers for disclosure of information, including within the framework of international legal assistance.

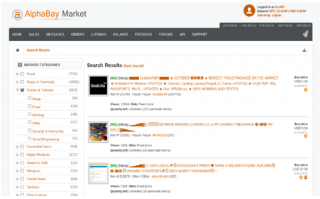
13. DarkNet Search

All data on the global network can be conditionally divided into three uneven segments – Surface Web (approximately 10% is publicly accessible and indexed by standard search engines), Deep Web (around 90%, consisting of webpages not indexed by these search engines and lacking links from surface resources) and Dark Web (DarkNet or shadow network, accessible only with the use of special software).

Unlike the *Surface Web* most *Deep Web* resources are not intended for wide public viewing (for example, certain sites of government institutions and commercial organizations, account pages on various web resources, cloud storage, subscription-based sites, databases, closed forums, catalogs, libraries, etc.). Typically, search engines can only find the homepage of such resources, not their content. Accessing these requires a standard browser and a direct link to the website (unlike conventional addressing, such a URL contains many random alphanumeric characters to complicate guessing), as well as a login and password (and sometimes IP or MAC address, if user verification is required).

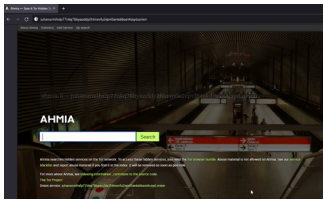
DarkNet (or onion network) consists of resources that use their own DNS (domains) and address space, such as .onion or .i2p top-level domains instead of conventional .com or .net. Connections among participants are encrypted and established using non-standard ports and protocols. As a result, this segment is inaccessible without browsers with special routing protocols (onion protocols) such as Tor, Onion Browser, OrNet. An alternative is the I2P (Invisible Internet Project), which is more secure and faster but less intuitive and popular. However, I2P provides access only to specific sites (so-called eepsites). The I2P network has its own catalog of sites, electronic libraries, and torrent trackers. In addition, there are gateways for accessing the I2P network directly from the Internet, created specifically for users who, for various reasons, cannot install I2P on their computers.

In essence, the DarkNet is a collection of anonymous computer networks designed to prevent tracking and control over information dissemination. This makes it a tool for bypassing restrictions, a confidential communication channel, or even a weapon for cybercriminals. For this reason, the DarkNet hosts resources from socio-political organizations; sites dedicated to investigations that are dangerous or banned from being published openly (e.g., the news site ProPublica), social networks and forums; email services; online libraries free of censorship; collections of interesting information; torrent trackers; and, most importantly, supplies, services, or content whose circulation is legally restricted or outright banned.



Here, various documents, bank cards, data leaks, account hacks, drugs and weapons, manuals or reference books, software, services, etc. are sold on trading platforms, forums or message boards. OSINT tools can also be used to try to de-anonymize attackers, collect their digital footprints (for example, by prolonged monitoring of network nodes, using social engineering methods, exploiting various vulnerabilities, using logins/passwords from the DarkNet in the public segment by inattention or accident, etc.) and prevent potential cyber or terrorist attacks, track illegal transactions. The currency used for illegal transactions is usually Bitcoin, which additionally ensures the anonymity of transactions. To date, one of the most effective ways to stop such illegal activities is to move the actions of criminals from the virtual world to the real world, for example, in the process of transferring drugs purchased on the shadow Internet.

Searching the DarkNet is quite labour-intensive due to *spam links* in search results or *changes in site addresses*. This segment is not designed to be a well-organized and indexed part of the network, as the main goal of most services is to remain hidden and accessible only to the "right" visitors. There are many different types of Dark Web search engines, and each has its own specialty. Navigation is facilitated through:



• *search engines* – Ahmia, Candle, Deep search, DuckDuckGo (the default search engine for the Tor browser), Excavator, Fess, GDark, Google.onion, Grams, HayStack, Kraken, Not Evil (ranks results), OnionLand Search, OnionSearch (GitHub, generates a file with results from different .onion search engines), Raklet, SearX (meta-search), Submarine, TorBot (GitHub, collects addresses and page titles with short descriptions), TORch (пsupports search operators and filters), TorDex, VigilantOnion (GitHub, oonion crawler with keyword search support) and others. Each engine produces different results for the same query, so it is better to use several;

• *link directories* – Daniel, Dark Catalog, Deep Links Dump, Deep Link Onion Directory, Hidden Links, Hidden Reviews, Hidden Wiki, Oneirun, OnionDir, Onion link list, Onion Links, Runion Wiki, The Dark Web Pug, Godnotaba and alternatives like darknet.wtf;

• *specific resources* – Archive.today (web archive), DarkNetLive (information about the DarkNet and its use), DarkVideo (YouTube analog); Dark Lair, Facebook (social network), Hidden Answers (forum); cryptocurrency platforms like GreenAddress, Onion Wallet, Smartmixer; email services like Mailpile, Mail2Tor, ProtonMail, SecMail, Sigaint; and libraries like Sci-Hub (scientific works), Just Another Library (literature).

Access to some sites, especially forums, may require a login and password, which can only be obtained through recommendations from other users, entry testing, or payment.

Thanks to traffic encryption and IP address masking technologies, the DarkNet provides a high level of anonymity for users. However, this does not mean that its use is entirely safe. There are numerous *risks and threats*, such as malware infections, phishing (stealing personal information, bank card/crypto wallet data, passwords, etc.), fraud, and tracking by both criminals and law enforcement.

To minimize these risks, comprehensive security measures must be taken when browsing and researching content, such as using separate physical devices or virtual machines, privacy-oriented operating systems, paid VPNs, fake identities, effective antivirus programs, , and, most importantly, common sense and caution.

Check the reputation of a website or forum before accessing it (e.g., via r/darkweb or r/TOR) and only visit links and resources you trust. Avoid suspicious or unreliable links that may lead to phishing sites or malware downloads. Do not click on pop-ups, advertisements, or suspicious requests.

Open downloaded files only after scanning, them, disconnecting from the internet, and preferably in an isolated software environment. Opening them while connected could potentially leak your real IP address and result in unforeseen consequences.

Thus, the DarkNet is a complex and ambiguous part of the Internet that can be used for both legal and illegal purposes. Understanding the principles of its functioning, as well as the methods of searching and analyzing the information obtained there, allows you to effectively use this knowledge in various areas, from crime investigation to human rights protection. It is important to always follow the rules of safe use of DarkNet and check the accuracy of the information received.

14. Useful Resources for OSINT Skills Improving

Ultimately, the most important thing is to keep updating your tools and staying informed about the latest developments in OSINT to remain aware of the newest trends and methods for searching and analyzing collected data. The following resources can help you:

• ***set ofinstruments/resourses*** – Advanced Search Tools, Analyst Research Tools, AsInt_Collection, Awesome OSINT, BBC Forensics Dashboard, Bellingcat's Online Investigation Toolkit, Commandergirl, CTI, Cyber Detective's website (or on the social network X/Twitter), DarkWeb, DeepWeb, Domainname-and-IP, EmailOsint, FBI-tools, Free OSINT and Online Research Resources, Free Osint Tools, IntelTechniques, MetaOSINT, NCSO, OSINT Essentials, OSINT Framework, osintframework.de, OSINTgeek Tools, OsintInception, OSINT Investigation Assistant, OSINT Research, OsintSmartFramework, OSINT Tool Comparison Table, OsintTools (by Molfar), OSINT tools (by Aware Online Academy), OSINT Web Resources, Osint4All, PhotoOsint, Search, SocialMedia, SPJ Toolbox (by Society of Professional Journalists), Technisette, The Hound, The Ultimate Osint Collection, Verification Toolset;

• ***specialized browsers/collections of bookmarks*** – Dark Web OSINT Bookmarks (for Tor), OSINT Bookmark Stack (for Chrome or Firefox);

• ***programs for visualizing research*** – Obsidian (guide available in Ukrainian), OSINTBuddy (GitHub, for visualization and finding starting points for further investigations, a free alternative to Maltego), TheBrain;

• ***practical cases/recommendations*** – Bellingcat's Guide, Dating apps and hook-up sites, Global Investigative Journalism Network, Online Research Cheat Sheets, OSINT Handbook 2020, OSINT Techniques, Technisette Tutorials, The Atypical OSINT Guide;

• ***training exercises*** – OSINT CTF/Challenges (catalog of resources for various quests, CTF games, web security training, and OSINT investigations), OSINT Exercises with Sofia Santos;

• ***thematic websites and Telegram channels*** – OSINT Team (a selection of YouTube channels, informational materials, blogs, podcasts, CTF games, hackathons, etc.); HackYourMom, InformNapalm, Molfar about OSINT, OsintFlow, OSINT Bees.

Виробничо-практичне видання

Зоренко Дмитро Сергійович
Кульчицька Людмила Олександрівна
Лех Роман Вікторович
Червяков Олександр Іванович

ВИКОРИСТАННЯ ІНСТРУМЕНТІВ ТА МЕТОДІВ OSINT ДЛЯ ОТРИМАННЯ ПОШУКОВОЇ ІНФОРМАЦІЇ

Практичний порадник
5-те видання, перероблене та доповнене

(англійською мовою)

ISBN 978-617-8130-72-5

Підписано до друку 21.03.25.
Формат 60×84 $^{1}/_{16}$. Папір офсетний.
Ум. друк. арк. 4,65. Гарнітура Times.
Наклад 50 прим.

Інститут Служби безпеки України
Національного юридичного університету імені Ярослава Мудрого
61002, м. Харків, вул. Мироносицька, 71,
телефон/факс: (057) 700-34-55, e-mail: ipuk@ssu.gov.ua

Видавець: Мірошниченко Олег Анатолійович
61002, м. Харків, вул. Дарвіна, 16, кв. 25.
Свідоцтво Державного комітету телебачення
і радіомовлення України
серія ДК № 5818 від 28.11.2017 р.
ел. пошта: merash@i.ua

Надруковано у друкарні ТОВ «Цифра Прінт».
Свідоцтво про Державну реєстрацію А01 № 432705 від 03.08.2009 р.
Адреса: 61166, м. Харків, вул. Данилевського, 30