Review study on Information Gathering Using OSINT for Forensic Investigations

Mekala Rahul Reddy[1], Bharadwaj M [2*]

[1]B.Sc. (Hons) Digital Forensic Science, Malla Reddy University, Hyderabad, India

[*2]Assistant Professor, Department of Digital Forensic Science, Malla Reddy University, Hyderabad, India

bharadwaj861@gmail.com[*2]

Abstract:

In the current era of Digital Forensics, evidence collection and gathering information related to the crime, victim and the attacker is crucial, these evidences can be available at the scene of crime or in the larger databases like the Internet. The data which is present in the internet about a person or an organization is found in the public records and databases. This collection of the data which is publicly available can be performed with the help of OSINT, Open-Source Intelligence, using OSINT we can systematically collect, analyze and interpret the public databases. OSINT tools are mostly found all over the internet with each domain focusing on a specific type of data. OSINT can also be performed using Kali Linux based commands for a simpler and broader user data search. This paper provides a review on the information gathering tools and software which are available for open source while analyzing the accuracy of the selected software along with its legal, technical and ethical considerations. This paper includes rigorous searching for the information over the internet along with evaluating its accuracy and relevance in accordance to the source. In the process of gathering information related to a specific person we find the person's digital footprint and their interaction with various platforms. This paper also highlights the future directions of OSINT, its credibility and reliability of the results.

Keywords: OSINT, OSINT Framework, Information Gathering, Digital Footprint, Reconnaissance, Digital Evidence

D. R. Reddy et al. (eds.), *Proceedings of the First International Conference on Advances in Forensics and Cyber Technologies (ICFACT 2025)*, Advances in Computer Science Research 127,
https://doi.org/10.2991/978-94-6239-610-4_24

1. Introduction:

With the expansion of the technology and the emergence of social media along with increase in databases over the internet, the ways to approach information gathering has taken a massive turn from its traditional methods and resources. In the current era with the usage of internet and being dependent on it for all the minor tasks, led to a significant amount of data being digitized on a daily basis through the creation of electronic registers known as the databases. For the legal and judicial process where evidence plays a key role in a criminal case to aid in solving the crimes not always the physical evidences can be helpful. For this investigator aim to collect as much information as possible with relevance to the crime. This evidence can be easily collected at the scene of crime where we mostly find all the physical evidences.

Due to technological development the methods of committing crimes have also faced a drastic turn of events, where the criminals are mostly committing online crimes which are also known as cybercrimes. For the investigation process of the cybercrimes to be followed collection of digital evidences is must. These evidences are sometimes available at the scene of crime, where the devices which are being used by the perpetrator provides most of the evidences related to the crime. Sometimes in some cases the investigators may not find all the evidences which direct the crime towards the suspect, so the investigators try to look for the evidences apart from the suspect interacted devices. This introduces forensic investigators with a concept known as OSINT. This is a collection open-source information collecting methods where investigators can use the software and tools which are available for open-source usage. OSINT commonly refers to the open-source intelligence where people use tools and software techniques to search for data which is publicly available for everyone to access. With the help of OSINT, it provides law enforcements and investigators with timeline reconstruction, mapping the criminal connections and their networks and provide evident user activity proofs of any specific individual without the need of acquiring physical access to the devices and hardware.

OSINT is defined as the systematic collection, processing, and analysis of publicly available data to generate actionable intelligence. OSINT gathers data from many different types of public sources, including social media, public databases, forums, government reports, and the dark web. [1]

In forensic science point of view, OSINT is not just defined as "searching the internet", but a systematic collection, and analysis of the Publicly Available Information (PAI) to generate evidences. In OSINT the raw data is considered as intelligence only after when it has been verified and interpreted. OSINT operates in various technological fields and it contains a wide spectrum of data sources, which includes:

- **Surface Web**, it consists of web which is easily visible over the internet without any minimal effort. It includes Social Media platforms, news channels, and public records. This is considered to be a small part of the internet where it is estimated to be less than 5%, it is familiar with all the internet users as it is easy to surf.
- **Deep Web**, is like a vast portion of the internet where it consists of non-indexed databases, academic repositories, and registration forums. It is not indexed by the standard search engines as it is designed for privacy, and to protect sensitive information of the users and organizations.
- **Dark Web,** is totally different from deep web and surface web, where it is a hidden part of the internet which is only accessible with specialized software like Tor Browser. This is mostly used for its illicit marketplaces and anonymized encrypted communication channels.

The integration of these sources with traditional forensic data has given rise to the concept of Digital Forensic Intelligence (DFINT), a holistic approach where OSINT is used to corroborate artifacts found on physical devices or to generate probable cause for search warrants. [2]

This paper presents a comprehensive review of the methodologies, tools and OSINT frameworks which helps in gathering of the information. This scope of the paper includes an analysis of both Surface web collection techniques along with deep and dark web. This paper mainly focuses on how OSINT can be integrated with traditional investigation process and the relevance of this gathered information.

To address the scope which is defined above, this review is guided by some core questions regarding the review. They are:

- How traditional forensic methodologies incorporate OSINT as an evidence source?
- What tools and techniques are available for collection and analysis of OSINT data?
- What are legal, ethical and technical challenges faced for OSINT evidence admissibility in court?

- What specific challenges and solutions exist for gathering intelligence from non-indexed web sources?

2. Conceptual Framework:

OSINT provides a systematic approach for developing meaningful insights from publicly available data. It is an integral part of any intelligence operation, whether online or offline. [3]

Though OSINT gained a widespread popularity and mainstream attention from public in early 21st century due to internet and social media, it is a well-established concept with a long history, During the World war era military organizations used local news communication channels, radio broadcasts and public records to gather the information of their enemy activities. Later during Cold War, many secret agencies monitored the actions of the enemy country and gained quite a lot of information to turn the war into their favor. The use of public resources and covert operations conducted by many secret organizations helped rival countries to gather information make themselves stand on top of nuclear race. These early initiatives laid the foundation for the current OSINT technology. Currently, 80–90% of intelligence activities in Western law enforcement and national agencies rely on OSINT. [4]

OSINT is currently used by many sectors and it has proven its efficiency in proving its capability. It is applied in many diverse sectors like cybersecurity, law enforcement agencies, private detective agencies, military organisations, journalism, and many more. OSINT proved its efficiency in providing required and reliable evidence by collecting the information from various online databases. With its importance in the current era, OSINT is not just considered as a passive library search, but as an active investigation method which sums all the acquired fragmented data into structured format. OSINT is used to scrape various blogs, forums, paste sites, and dark web marketplaces to track leaked credentials, security threats, zero-day exploits, and potential cyberattack discussions. [5] In addition to this, OSINT tools are also used to track social media profiles who are suspected to be fake, criminal networks, terrorism, online scams, espionages.

3. Types of Open Sources:

OSINT has evolved into many categories which provide various types of data which can be quite helpful in any investigations. OSINT, consists of a collection of various sources, where it is primarily categorised into five domains in relevance to forensic investigations namely- SOCMINT, GEOINT, Surface Web & Public Records, Deep Web Sources, Dark Web Sources.

1. **Social Media Intelligence (SOCMINT):**
    - **Definition-** SOCMINT or Social Media Intelligence is a subdiscipline of OSINT. It collects and analyses publicly available data from social media platforms such as Facebook, Twitter (X), Instagram, and LinkedIn. SOCMINT is used by governments, law enforcement agencies, and business companies to assess public opinion and identify risks, trends, or threats. [6]
    - **Forensic Significance-** It helps in providing behavioural patterns, social connections, emotional attachments, and temporal data of user or suspect which are believed to be crucial for establishing timelines for activities. [7]

2. **Geospatial Intelligence (GEOINT):**
    - **Definition-** GEOINT or Geospatial Intelligence is a branch of intelligence that collects, analyses, and interprets geographic information to support military, security, and humanitarian efforts. Unlike HUMINT or SIGINT, it relies on visual and spatial analysis to identify patterns, track threats, and improve awareness. GEOINT uses satellite imagery, aerial reconnaissance, and mapping technology to provide critical insights into terrain, infrastructure, and human activities. [8]
    - **Forensic Significance-** It is mostly used to verify the alibis provided by suspects, locate suspects activity, and validate physical events with the acquired digital events. [9]
3. **Surface Web & Public Records:**
    - **Definition-** Surface web is the visible part of the internet where we can find data through different search engines available for us. The indexed content accessible via standard search engines like Google, Bing, Yahoo, DuckDuckGo, Brave. This includes corporate registries, court records, domain registration data, news archives, property records, government statistics.
    - **Forensic Significance-** It is quite essential for verifying identities, performing background checks, and corporate due diligence. [10]
4. **Deep Web Sources:**
    - **Definition-** Deep Web OSINT involves gathering intelligence from non-indexed parts of internet, like behind login portals, leaked information and

public filings. This is mostly used to map networks using both the publicly accessible information and the deep web information. It provides data obtained from behind logins, like publicly available forums that require subscription, corporate records and sanctions lists. In these instances, the information is intended and available for public consumption but requires user authentication, unlike the surface web. [11]

- **Forensic Significance-** It often contains data which is of high fidelity than the data present in surface web. It includes technical documentation and specialized community discussions. [12]

5. **Dark Web Sources:**
    - **Definition-** Dark Web Intelligence (DARKINT) is a specialized subset of OSINT which focuses on gathering information from the hidden, encrypted portions of the internet not indexed by standard search engines. It is different from traditional OSINT as it uses surface and deep web for its data, whereas this deals with specialized tools and techniques to access and analyse information for cybersecurity, law enforcement agencies, and threat intelligence purposes. [13]
    - **Forensic Significance-** DARKINT is critical for investigating cybercrime, illicit marketplaces, data breaches, and threat actor communications. While this data is Hidden, it is considered as Open Source if it is accessible to the general public without hacking any domains or databases. [10,14]

4. OSINT relevance in Digital Forensics:

The integration of OSINT into digital forensics represents a shift from a device-centric to a data-centric investigation model. It is a rather complex multidisciplinary task to do digital evidence search, collection, extraction and analyzing, because a person doing that should have thorough forensic, criminal proceedings, and technical knowledge [15]. The relevance of OSINT in digital forensics is threefold, where they are;

- **Enrichment & Contextualization:** Traditional forensics might yield a phone number from an acquired device in scene of crime. OSINT enriches this artifact by linking the number to a social media profile, registration details, physical address, which aids in instantly widening the investigative scope [16].
- **Pre-Warrant Investigation:** In many jurisdictions, law enforcement agencies cannot seize devices without a warrant and a proper cause. OSINT allows investigators to

build this cause through using non-intrusive means, identifying targets, and their networks before any physical intervention occurs. [17]

- **Attribution in Cybercrimes:** In cases of hacking or ransomware, where the suspect is impossible to physically reach, OSINT can be used to gather the information regarding the attack and the suspect, and sometimes it can be the only means of accrediting an attack to a specific individual or group. [18]

5. OSINT Methodology and Tools Landscape:

OSINT methodology and tools can be organized into categories which mirrors the investigation process, from the detection to automation. There are several categories in this, and they are as;

- **Search engine OSINT:**

  Search engines are considered as the first layer in any OSINT investigations. These are used with addition of advanced operators rather than just using plain queries for search [19]. Google Dorking is one of the techniques where investigators use google search bar to search for information by using specific operators and terms. It helps investigators to uncover exposed directories, misconfigured services, login portals, cached data, leaked documents that usually doesn't appear by performing normal searches. [20]

  Operators: Google Dorking can be performed by using operators like; site: filetype: inurl: intitle: all for specific targeted discovery of documents, device interfaces and credentials. [20]

- **SOCMINT:**

  SOCMINT mainly focuses on extracting and analyzing information from various platforms to understand identities, relationships, connections and many more. It helps in using user generated content for constructing timelines, observe social networks.

  ### Tools:

  Maltego- Maltego runs preliminary OSINT searches on suspects and threats. It accesses data from social media platforms, dark web sources, breached database, and identity databases directly from the web without compromising security. It is command-line based tools which can be operated in Kali Linux. It helps in pulling out posts, profiles, followers, mentions, and other entities from various social media platforms.[21]

SpiderFoot- SpiderFoot is an open-source intelligence (OSINT) tool. It integrates with every data source available and utilises a range of methods for data analysis, making the data easy to navigate. It has an embedded web-server for providing a clean and intuitive web-based interface but can also be used completely via the command-line. It's written in Python 3 and MIT-licensed. It helps in enumerating social accounts, find aliases, and pivot between usernames, emails, and domains. [22]

Sherlock- Identifies social media profiles on over 300+ platforms based on a given username. It runs on all platforms like Windows, macOS, Linux. Sherlock hunt down social media accounts by usernames across social networks. It is an OSINT command-line utility used to find usernames across hundreds of social networking sites. It helps investigators and security professionals determine a person's online presence by checking if a specific username is registered on various platforms.[23]

- **Kali Linux Based OSINT Tools:**

  Kali Linux remains a popular distribution for security and forensic professionals because it contains an excessive suite of OSINT and reconnaissance tools. Commonly used Kali based tools for evidence-oriented OSINT include; **theHarvester-** for enumerating emails, domains, and hosts from multiple search engines; **SpiderFoot-** for automated multisource scanning; **Recon-ng-** for modular web reconnaissance.
  Semi-automated frameworks like snOint, which is a package in Kali, provide a structured way to enumerate attack surfaces and map the public information into unified datasets useful for the further investigations.

6. OSINT in Forensic Investigations:

Forensic investigations carried out through using OSINT follows as specific structured workflow that mirrors classical digital forensics, but is adapted to volatile, online sources. Each phaser carried out in the OSINT is always documented to preserve integrity and the chain of custody. In OSINT all the methods should be followed in a specific manner, where the order should be maintained for the investigation process. The Intelligence Cycle as defined by the Office of the Director of National Intelligence is as shown in figure 1. The process is step-by-step and reiterative. The evaluation stage informs the parameters implemented during the planning of subsequent investigations.[24]

The steps which are needed to maintained are;

Planning and Objective Definition: Before starting the OSINT forensic investigations, investigators should define clear intelligence and the evidential objectives before trying to touch any data available online.[25] Investigators should select the required tools, software, platforms and the data sources in advance and set timelines, responsibilities and legal constraints like the warrants, policies and jurisdictional limits as a part of the formal collection plan.[10]

Data Discovery: Investigators conduct high-level reconnaissance to understand the target's Digital Footprint like, usernames, domains, email addresses, infrastructure, social platforms and user preferred communities. Investigators use search engines and automated OSINT platforms to identify the relevant profiles, mentions, artifacts and infrastructure without collecting evidences in-depth.

Collection and Preservation: After the reconnaissance, when the investigators find relevant sources are identified, collect data in a forensically aware manner, which is to capture the URLs as a whole, timestamps, page contents, media and the contextual information using tools that preserve original source material. Investigators immediately apply preservation practices familiar from digital forensics, like hashing exports, keeping read only originals, keeping regular backups and documenting every action in an evidence log, known as chain of custody.

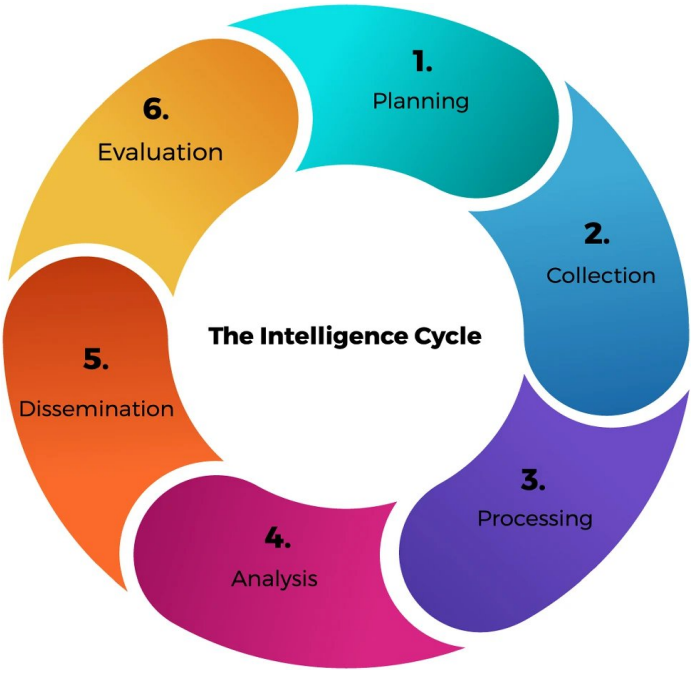

Fig 1 The Intelligence Cycle as defined by the Office of the Director of National Intelligence. The process is step-by-step and reiterative. The evaluation stage informs the parameters implemented during the planning of subsequent investigations.[24]

Verification and Correlation: For each and every OSINT tool present in the Internet, investigators treat each OSINT item as lead evidence until verified. It helps in corroborating identities of the suspects, locations and timelines across multiple independent sources like social post and public records infrastructure data. Investigators evaluate reliability by considering source history, consistency across platforms and technical indicators like metadata, headers, network data and document analytic methods for later analysis.

Reporting and Evidence Presentation: After the evidence has been collected and processed and analyzed investigators need to present their findings in a structured report to separate raw artifacts, analytic reasoning and conclusions with clear references back to preserved original select hash values, filenames and capture times. Investigators also need to include the methodology they followed, the tools which have been used, the limitations in their work and any uncertainties which are faced so that another examiner or the expert in a court can

understand and reproduce or challenge the process which has been done. It helps in satisfying the evidence and it also maintains a chain of custody. Creating report which has no errors in it helps in maintaining the integrity.

Chain of Custody Concerns: Investigators need to make a copy of every transfer, access to OSINT tool derived evidence like screenshots, exports, archives, must be logged with date, time and a handler, and the purposes. Investigators should follow principles from ACPO, NIST and similar standards. Investigators while working on the crop is should be used which should be used for analysis while original capture should remain untouched and securely stored It helps in rearing best practices and device forensics and never the original data should be worked on so that the integrity can be easily maintained.

7. Applications in Forensic Casework:

OSINT has moved from an exterior to central pillar in many types of forensic investigations. In many domains, it helped a lot like real cases show how open-source data help to identify and reconstruct events, and it helps to strengthen evidences collected. Some of the domains of it are;

- **Cyberstalking and Harassment:**
  In many of the documented cyberstalking and harassment cases, investigators have used OSINT to record the offenders online identities, find linked anonymous accounts to real profiles and document patterns of abusive messages across platforms, Investigators also use social media timelines, reused usernames and Geotagged posts to correlated and show closeness to the victim, growth overtime and breach of restrictive guidelines, providing courts with a structured evidential record rather than isolated screenshots.[27]

- **Fraud and Financial Crimes:**
  Many case-based research on OSINT in cyberespionage and financial crimes show how investigators correlate company databases, leaked credential dumps, domain records and blockchain traces to expose shell companies, mule accounts and money laundering routes. In APT and financial fraud operators, open-source malware reports, passive DNS data and cryptocurrency tracking allowed analysts to attribute infrastructure, follow illegal funds and support accusations by linking public technical indicators to specific groups and individuals.[28]

- **Missing Persons:**
  Many public initiatives such as Trace Labs CTFs and documented case studies from investigators show how open sources help locate leads in real missing person investigations, Forensic investigators have combined previous social media posts, alternative account names, forum activities, breached credentials and public records to identify new contacts, their last known online activity and prior addresses materially expanding the lead set available to the law enforcement.

- **Addressing Cybercrime:**
  OSINT is increasingly recognized as a critical tool in cybercrime investigations. It provides investigators with the ability to expand their limited technical artifacts into comprehensive profiles, infrastructure and criminal methodologies. Academic research demonstrate that OSINT enables the mapping of cybercriminal networks and support acknowledgment to the analysis of open-sourced data such as domain registration, social media activity and infrastructure reuse. For example, a study on cyberthreat attribution highlights how awesome combined with technical forensics allows analysts to think desperate additions into larger operations where correlating malware signatures, IP addresses and domain data found in open repositories and public reports. [29]

8. Challenges and Limitations:

OSINT gathers actionable insights from the vast and unstructured nature of publicly available data. Despite its enormous opportunities it has significant challenges that limit its effectiveness. When OSINT practitioners gather data from search engines, social media, forums, newspapers or other platforms they often encounter critical challenges such as fragmented data, anonymous sources or deliberately misleading descriptions. To mitigate these challenges investigators must focus on legal and privacy regulations, data bias issues, AI tools or techniques and deepfake technologies these factors significantly impact the accuracy reliability and ethical use of open-source intelligence [3].

- **Data Overload:** OSINT collects and analyses vast amounts of data from multiple sources, including social media, news outlets, forums, and government records. The primary challenge of OSINT is the sheer volume of data. Such vast amounts of data are hard for humans to interpret. Every day, approximately 500 million tweets are posted,

over 4.75 billion items are shared on Facebook, more than 500 hours of video are uploaded to YouTube per minute, and Google handles over 99,000 search queries per second [31,32].

- **Data Accuracy and Reliability:** OSINT relies on publicly available data, and not all public data is reliable or accurate, so it can be a double-edged sword for OSINT analysts. Misinformation, disinformation, and propaganda can distort the intelligence-gathering process. A high-profile example of this challenge was during the 2016 US presidential election when the Russian disinformation campaign on social media significantly influenced public opinion. Inaccurate data undermines OSINT's reliability and can easily mislead analysts and decision-makers [33]. Open sources are mostly uncontrolled; so sometimes the information maybe outdated, incomplete or intentionally planted to mislead the investigation process [34].
- **Ethical and Legal concerns:** in OSINT investigations, investigators are often faced with problems like ethical concerns which conflict with individuals' privacy rights, legal frameworks and ethical boundaries.[35] The primary reason for this is the presence of a line between what data is publicly available and what data violates the privacy is not always clear. Even if data is available publicly, collecting and processing Personal Identifiable Information (PII) of individuals requires a warrant. Collecting others personal information without a proper cause violates individuals' privacy, and sometimes this leads to the evidence to be considered inadmissible in the court of law, which makes the investigator to face with the consequences according to law.[36]
- **Rapidly Changing Platforms:** Platforms like social media frequently alter their algorithms, remove old features and add new features often. Some social media platforms take anti-scraping measures, to secure their user data from reaching public without a proper access. With this rapid change investigators need to be frequently updated with the technological change and they need to constantly retrain and adapt their methods to progress in investigations.

9. Comparative Review of Existing Literature:

Review of the current academics revels a rapid changes and transitions. While the earlier studies focused mostly on the manual extraction of the data, now with the technological development the era has changed investigators use OSINT for online information gathering using public databases, integrating AI in search, and legal standardization.

Table1: Comparative review of literature

S.N.	Paper Title	Authors & Year	Country/ Region	Work Focus	Conceptual Framework	Research Gap
1	OSINT Tools & Resources Handbook 2020	Aleksandra Bielska, Noa Rebecca Kurz, Yves Baumgartner, Vytenis Benetis (2020)	Switzerland	Comprehensive OSINT toolkit; with 70+ categories covering search engines, social media, eople investigations, legal resources.	Categorized resource taxonomy; intelligence application domains classification.	Limited evaluation of tool effectiveness; no guidance on tools mentioned, insufficient coverage of emerging platforms.
2	How Dark Web Monitoring Can Be Used For OSINT & Investigation	Allanoud Alquwayzani, Rawabi Aldossri, M.M. Hafizur Rahman (2023)	Saudi Arabia (King Faisal University)	Dark Web monitoring using OSINT; illegal activity tracking; AI/ML threat detection.	PRISMA systematic review methodology; dark web architecture and technologies analysis; AI enhanced monitoring framework.	Insufficient advanced AI techniques; limited ethical/legal framework guidance; knowledge gap in dark web forensic standards.
3	Application of OSINT Methods in Ensuring Cybersecurity	Sabina Szymoniak, Kacper Foks, Aleksandra Pyrkosz Dziubczyk (2024)	Poland	Defensive & offensive OSINT in cybersecurity; threat detection; vulnerability identification.	OSINT dual use framework; intelligence cycle integration.	Information overload; unreliable data sources; lacking ethical guidelines; malicious actor misuse of OSINT; privacy concerns.

4	Open sources Intelligence Opportunities and Challenges – A Review	Sabina Szymoniak, Kacper Foks (2024)	Poland (Czestochowa University of Technology)	OSINT tools, techniques, applications; privacy implications; metadata analysis; social media forensics.	OSINT evolution framework; social media analysis methodology; privacy-security balance model; content verification approach.	Limited systematic tools evaluation; insufficient data quality assurance; emerging deep fakes/ anonymization challenges not adequately addressed.
5	A Systematic Review on Research Utilizing AI for OSINT Applications	Thomas Oakley Browne, Mohammad Abedin, Mohammad Jabed Morshed Chowdhury (2024)	Australia (La Trobe University)	AI/ML algorithms in OSINT; intelligence cycle automation; threat detection systems; machine learning applications.	PRISMA systematic review; AI across intelligence cycle phases (planning, collection, processing, analysis, dissemination); RNN, LSTM, CNN,NLP applications .	AI-OSINT tool integration gaps; limited penetration testing models; underutilized alternate data sources; weak dissemination functionality .
6	OSINT as an Investigative Tool: Problems	Kateryna Latysh,9 Yevheniia Demidova, Mariieta	Ukraine (Yaroslav Mudryi National Law	OSINT in war crimes investigations; digital evidence standardization;	Berkeley Protocal Framework; Leiden Guidelines	Unified international standards lacking; legislative unpreparedness for

	of Collection and Standardization	Kapustina(2023)	University)	investigative protocols; international cooperation.	Integration; Rome Statue compliance; multidisciplinary approach (Legal, technical, procedural).	digital evidence; GDPR compatibility gaps; NGO involvement protocols undefined.
7	The Art of Open Source Intelligence: Addressing Cybercrime, Opportunities, and Challenges	MD Sazibur Rahman (2023)	China (Xi 'AN University of Posts & Telecommunications	OSINT principles, deployment strategies; cybercrime prevention; legal/ethical/technical considerations.	Case Study methodology; intelligence gathering methods (HUMINT, SIGINT, GEOINT, MASINT, FININT, SOCMINT); Threat detection framework.	Data accuracy validation challenges; deepfake proliferation; legal ambiguities across jurisdictions; privacy-security balance complexity.
8	Role of Multimedia Information Retrieval In Digital Forensic Investigations Using OSINT	Amr Adel, Brian Cusack (2020)	New Zealand (Auckland University of Technology)	Multimedia analysis in digital forensics; data mining; link analysis; criminal intelligence building.	OSINT Forensics framework; data lifecycle management; entity identification; five digital forensics challenges model	Five major challenges; complexity, diversity, consistency/correlation, volume management, unified time lining; automation tool insufficiency.

					(complexity, diversity, consistency, volume, time lining).	
9	OSINT by Crowdsourcing: A Theoretical Model for Online Child Abuse Investigations	Kemal Veli Acar (2018)	Turkey (Turkish National Police)	Crowdsourced OSINT for digital forensics backlogs; child abuse investigations; volunteer co-ordinations.	Crowdsourcing framework with OSINT; Technical, legal, organizational aspects; volunteer management model .	Evidence backlogs causing years of delays; insufficient LEA resources; automated methods inadequate for detailed investigation; human-in-loop approach necessity.
10	Investigations Methods for Forensic Analysis of Social Media Data to Support Criminal Investigations	Muhammad Arshad, Arshad Ashfaq Ahmad, Choo Won Onn, Emmanuel Arko Sam (2025)	Multi-Institutional (Ireland, Pakistan, Malaysia)	AI/ML Driven social media forensics; privacy compliance; data integrity challenges.	AI/ML integration framework; text mining, network analysis, facial recognition, tampering detection; legal/ethical compliance emphasis.	Data deletion/editing challenges; privacy law restrictions (GDPR, CCPA); algorithmic bias in facial recognition; insufficient training datasets; evidence admissibility concerns.

10. Ethical and Legal Considerations:

OSINT investigations mostly rely on publicly available data; they operate in a complex legal framework where the boundaries between public and private is often blurred. Investigators should stick to strict guidelines to ensure their findings are legal, ethical and admissible in court.

Admissibility of OSINT in Court:

For the OSINT evidences to be used as evidence in legal proceedings, OSINT data must satisfy three main criteria:

- **Authentication Precedents:** In United States v, Browne (2016), the court established that social media evidences must be properly authenticated to be admissible. This requires corroborating metadata such as timestamps and IP addresses rather than just providing screenshots, which are easier to forge. [7]
- **Chain of Custody:** Investigators must maintain an undisputable record of the data collection process carried out by them.
  Hashing-Cryptographic hashes like SHA256, MD5; should be generated for every captured file to prove it has not been altered since collection, and it is maintaining its integrity.[7]
  Blockchain Logging- Developing practices use blockchain technology to create tamper proof timestamped logs of evidence collection, ensuring the integrity of data path.[7]
- **Metadata Validation:** Geotags and EXIF data are crucial for verifying the time and location of digital evidences, and separating the authentic content from fabricated media.[12]

Privacy Implications:

While OSINT investigations use publicly available information, the collection of this data can cause violation of individuals privacy rights.

- **Data Protection Laws:** Regulations like **GDPR** (General Data Protection Regulation) and **CCPA** (California Consumer Privacy Act) imposes a strict boundary on collecting

and processing personal data. Even when the data is publicly available, processing it for profiling without any legal cause may violate the individual's privacy rights. [7]

- **Stalking Boundary:** There is a fine line between legal investigations and cyberstalking in case of OSINT investigations. Monitoring an individual's online activity excessively without any clear cause can be legally considered as harassment or stalking, which is a punishable offence.[7]

11. Future Directions:

Several trends point towards the future evolution of OSINT in digital Forensics and law enforcement. Research on OSINT frameworks proposes a unified systems that integrate social, network, telecom, blockchain, leak, and identity-based intelligence into convergent architectures for multidomain cybercrime investigations. Digital forensic readiness concepts, such as centralized intelligence repositories and standardized data models aim to ensure that OSINT data is captured and structured in many ways that facilitate both operational use and forensic examination.[40]

Standardization and Best Practice Guidelines: There is a need for formal standardization of OSINT methodologies, data models and reporting standards suitable for forensic contexts. Future work should develop:

- **Unified Data Models:** Standard representation for representing OSINT findings enabling interoperability between tools and organizations. [40]
- **Validation Frameworks:** Standardized protocols for verifying OSINT accuracy, assessing tool reliability and benchmarking platform performance must be carried out by the investigators.

12. Conclusion:

Open-Source Intelligence has established as a critical and increasingly essential component of modern digital forensics, enabling investigators to contextualize local device evidence within the broader digital ecosystem of social media, public records and technical infrastructure. OSINT tools spans from freely available Kali Linux tools to commercial platforms. These made investigators to use open-source intelligence while using methodological, legal, and ethical complications that investigators must solve. The integration of OSINT with traditional digital forensic methods has introduced new frameworks like DFINT, which propose semi-

automated convergence of local device data with global open-source intelligence to amplify investigative approaches. This paper discusses the core concepts, tools, techniques, and practical implementations of Open-Source Intelligence, highlighting its critical role in detecting, investigating, and combating cybercrime. OSINT investigations not only strengthen the forensic investigations but also provides opportunities for threat assessment. However, despite its strengths it faces limitations such as data overload, data manipulation, outdated information, unreliable sources, privacy restrictions, legal and ethical concerns. In conclusion OSINT serves as a powerful blend of art and science for information gathering, when it is ethically used investigators can find new leads in their cases, which aids in solving.

References:

[1] Open Source Intelligence Investigation. (2016). In B. Akhgar, P. S. Bayerl, & F. Sampson (Eds.), *Advanced Sciences and Technologies for Security Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-319-47671-1

[2] Quick, D., & Choo, K.-K. R. (2018). Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINT+OSINT): A timely and cohesive mix. *Future Generation Computer Systems*, *78*, 558–567. https://doi.org/10.1016/j.future.2016.12.032

[3] Rahman, M. S. (2025). The Art of Open Source Intelligence (OSINT): Addressing Cybercrime, Opportunities, and Challenges. *Journal of Computer Science and Technology*.

[4] Ghioni, R., Taddeo, M., & Floridi, L. (2023). Open source intelligence and AI: a systematic review of the GELSI literature. AI & Society, 39(4), 1827–1842. https://doi.org/10.1007/s00146-023-01628-x

[5] Browne, T.O., Abedin, M. & Chowdhury, M.J.M. A systematic review on research utilising artificial intelligence for open source intelligence (OSINT) applications. Int. J. Inf. Secur. 23, 2911–2938 (2024). https://doi.org/10.1007/s10207-024-00868-2

[6] Macêdo, A., Peotta, L., & Gomes, F. (2023). A Review of the Intersection Techniques on Humint and Osint. International Journal on Cybernetics & Informatics (IJCI), 12(1), 53

[7] Arshad, M., Ahmad, A., Onn, C. W., & Sam, E. A. (2025). *Investigating methods for forensic analysis of social media data to support criminal investigations*. *Frontiers in Computer Science, 7*, Article 1566513. https://doi.org/10.3389/fcomp.2025.1566513

[8] Heningtiyas, H., & Supriyadi, A. A. (2024). Study of the implementation of geoint and remote sensing in climate change. Remote Sensing Technology in Defense and Environment, 1(2), 45-55.

[9] Dekens, N. (2025, June 9). *OSINT techniques: Complete list of expert tactics for investigators.* ShadowDragon. https://shadowdragon.io/blog/osint-technique

[10] Neotas. (n.d.). *OSINT tools and techniques: A comprehensive guide on open source intelligence tools and techniques for risk, compliance and cyber investigations*. Neotas. https://www.neotas.com/osint-tools-and-techniques/

[11] Clarke, S. (2021, December 2). *Dark web vs deep web vs surface web: And how to incorporate each within OSINT investigations.* Blackdot Solutions. https://blackdotsolutions.com/blog/dark-web-vs-deep-web#:~:text=The%20deep%20web%20in%20OSINT,far%20more%20difficult%20to%20navigate.

[12] Szymoniak, S., & Foks, K. (2024). *Open Source Intelligence opportunities and challenges – a review*. *Advances in Science and Technology Research Journal, 18*(3), 123–139. https://doi.org/10.12913/22998624/186036

[13] OSINT Industries Team. (2025, November 18). *OSINT basics: What is Dark Web Intelligence (DARKInt)?* OSINT.Industries. https://www.osint.industries/post/what-is-dark-web-intelligence-darkint-beginners-guide#:~:text=In%20short%2C%20Dark%20Web%20Intelligence,the%20web%20will%20never%20see

[14] Alquwayzani, A., Aldossri, R., & Rahman, M. M. H. (2023). *How dark web monitoring can be used for OSINT and investigations. Journal of Theoretical and Applied Information Technology, 101*(10), 3838–3849. https://www.jatit.org/volumes/Vol101No10/15Vol101No10.pdf

[15] Latysh, K. (n.d.). *Concept of digital forensics and open data source information in the investigation process*. Vilnius University; Yaroslav Mudryi National Law University.

[16] Quick, D., & Choo, K.-K. R. (2018). *Digital forensic intelligence: Data subsets and open source intelligence (DFINT + OSINT): A timely and cohesive mix*. *Future Generation Computer Systems*, **78**, 558–567. https://www.sciencedirect.com/science/article/abs/pii/S0167739X16308639

[17] Social Links. (2025, July 10). *OSINT in criminal investigations: Methodologies and best practices.* Social Links Blog. https://blog.sociallinks.io/osint-in-criminal-investigations-methodologies-and-best-practices/

[18] Social Links. (2025, June 17). *SOCMINT: A pillar of the modern investigation and next-gen security.* Social Links Blog. https://blog.sociallinks.io/socmint-a-pillar-of-the-modern-investigation-and-next-gen-security/

[19] OSINT Industries Team. (2025, October 10). *OSINT basics: Going beyond Google with Bing and Yahoo dorking.* OSINT.Industries. https://www.osint.industries/post/osint-basics-going-beyond-google-with-bing-and-yahoo-dorking#:~:text=OSINT%20Basics%3A%20Going%20Beyond%20Google%20with%20Bing%20and%20Yahoo%20Dorking

[20] Makhrov, V. (2025, June 16). *Google dorking in cybersecurity: Techniques for OSINT & pentesting.* Netlas Blog. https://netlas.io/blog/google_dorking_in_cybersecurity/

[21] Paterva Pty Ltd. (n.d.). *Maltego — Interactive OSINT & forensics application.* https://www.maltego.com/

[22] Smicallef, S. (n.d.). *SpiderFoot* [Computer software]. GitHub. https://github.com/smicallef/spiderfoot

[23] Offensive Security. (n.d.). *Sherlock — Find usernames across social networks* [Software]. Kali Linux. https://www.kali.org/tools/sherlock/#:~:text=Find%20usernames%20across%20social%20networks,Across%20Social%20Networks%20(Version%200.15

[24] Bielska, A., Kurz, N. R., Baumgartner, Y., & Benetis, V. (2020). *Open source intelligence tools and resources handbook*. i-Intelligence GmbH.

[25] SOSIntel. (2025, February 27). *OSINT essentials: Planning, recording and evaluating intelligence.* SOSIntel. https://sosintel.co.uk/osint-essentials-planning-recording-and-evaluating-intelligence/

[26] Felix, A. (2025, February 21). *OSINT unveiled: The ultimate guide to open source intelligence in 2025.* Medium. https://medium.com/@Techwithhearts/osint-unveiled-the-ultimate-guide-to-open-source-intelligence-in-2025-9c5036a6670f

[27] Shodh Forensic. (2025, July 30). *Social media forensics: Foundations, technical frameworks and emerging challenges.* Shodh Forensic. https://shodhforensic.com/2025/07/30/social-media-forensics-foundations-technical-frameworks-and-emerging-challenges/

[28] Prasad, N., Diro, A., Warren, M., & Fernando, M. (2025). *A survey of cyber threat attribution: Challenges, techniques, and future directions. Computers & Security, 157*, 104606. https://doi.org/10.1016/j.cose.2025.104606

[29] Swate, C., Sithungu, S., & Lebea, K. (2024). *An analysis of cyberwarfare attribution techniques and challenges*. In *Proceedings of the 23rd European Conference on Cyber Warfare and Security (ECCWS 2024)*. Academic Conferences International.

[30] Rahman, M. S. (n.d.). *The art of open source intelligence (OSINT): Addressing cybercrime, opportunities, and challenges*. School of Computer Science and Technology, Xi'an University of Posts & Telecommunications.

[31] Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2020b). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. Expert Systems With Applications, 164, 114006. https://doi.org/10.1016/j.eswa.2020.114006

[32] Springer, S., Strzelecki, A., & Zieger, M. (2023). Maximum generable interest: A universal standard for Google Trends search queries. Healthcare Analytics, 3, 100158. https://doi.org/10.1016/j.health.2023.100158

[33] Klouček, T., Lagner, O., & Šímová, P. (2015). How does data accuracy influence the reliability of digital viewshed models? A case study with wind turbines. Applied Geography, 64, 46–54. https://doi.org/10.1016/j.apgeog.2015.09.005

[34] OSINT Team. (2025, April 22). *Assessing paid OSINT tools: Benefits, limitations, and considerations.* OSINT Team Blog. https://osintteam.blog/assessing-paid-osint-tools-benefits-limitations-and-considerations-98948de9ac02

[35] Riebe, T., Biselli, T., Kaufhold, M., & Reuter, C. (2023). Privacy Concerns and acceptance Factors of OSINT for Cybersecurity: A Representative survey. Proceedings on Privacy Enhancing Technologies, 2023(1), 477–493. https://doi.org/10.56553/popets-2023-0028

[36] Goodman Law Firm. (2025). *The legal implications of using OSINT (Open Source Intelligence).* Ernest Goodman Law Firm. https://ernestgoodmanlawfirm.com/the-legal-implications-of-using-osint-open-source-intelligence/

[37] Latysh, K. V., Demidova, Y. Y., & Kapustina, M. V. (2023). *OSINT as an investigative tool: Problems of collection and standardization*. In *Digital transformation and legal regulation* (pp. 131–149). https://doi.org/10.30525/978-9934-26-400-9-6

[38] Alquwayzani, A., Aldossri, R., & Rahman, M. M. H. (2023). *How dark web monitoring can be used for OSINT and investigations*. *Journal of Theoretical and Applied Information Technology, 101*(10), 3838–3849.

[39] Açar, K. V. (2018). OSINT by crowdsourcing: A theoretical model for online child abuse investigations. International Journal of Cyber Criminology, 12(1), 206–229. https://doi.org/10.5281/zenodo.1467897

[40] Soni, P., & Mathur, N. (2025). *A unified OSINT framework for multi-domain cybercrime investigation. Journal of Emerging Technologies and Innovative Research (JETIR), 12*(12), d201–d205. https://www.jetir.org/papers/JETIR2512326.pdf