This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.


doi: 10.11610/isij.5302

Available at: [https://doi.org/10.11610/isij.5302](https://doi.org/10.11610/isij.5302)

**CC BY-NC 4.0**
OSINT on the Dark Web: Child Abuse Material Investigations

Jyri Rajamäki (✉), Iiro Lahti, and Johanna Parviainen
Laurea Leppävaara, Laurea University of Applied Sciences, Espoo, Finland
https://www.laurea.fi

ABSTRACT:
The Dark Web allows users to hide their identity while browsing or sending information, providing an ideal environment for transferring information, goods, and services with potentially illegal intentions. Therefore, Law Enforcement Agencies (LEAs) are interested in Open Source INTelligence (OSINT) on the Dark Web. LEAs need appropriate techniques to find darknet sites used by criminals. This article examines online child sexual exploitation and the various OSINT automation tools that can be exploited on the Dark Web. Additionally, we consider OSINT on the Dark Web, paying attention to the challenges LEAs face when investigating crimes related to child abuse material (CAM). The biggest challenges are related to data storage and the criminal investigation itself. CAM may not be recorded or examined except by an LEA officer specifically designated and trained for this purpose. The study examines how OSINT could be implemented without exposing researchers to the contents of CAM. The method could be to focus the inquiry on already known links and sites. This has challenges, but a bigger number of LEAs could carry out such an inquiry, and the storage of such data would not be illegal.

ARTICLE INFO:
RECEIVED: 04 MAY 2022
REVISED: 15 AUG 2022
ONLINE: 19 SEP 2022

KEYWORDS:
OSINT, dark web, child abuse material, CAM, investigation, cybercrime

Creative Commons BY-NC 4.0

Introduction
Cybercrime is the priority of the European Multidisciplinary Platform Against Criminal Threats (EMPACT), and “Combating child sexual abuse & exploitation (CSA/CSE)” is EMPACT’s sub-priority. According to the most recent reports of
the Internet Organised Crime Threat Assessment (IOCTA), cybercrime is becoming more aggressive and involves various forms, such as high-tech crimes, data breaches, and sexual extortion. Cybercrime is a growing problem for countries, including the EU Member States. Data is a key target for cybercriminals, the number, and frequency of data breaches are on the rise, and this, in turn, is leading to more cases of fraud and extortion. Online child sexual exploitation is a constantly evolving phenomenon and is shaped by digital technology evolution. Mobile connectivity, growing internet coverage in developing countries, and the development of pay-as-you-go streaming solutions, which provide a high degree of anonymity to the viewer, are furthering the trend in the commercial live-streaming of child sexual abuse. EUROPOL/EC3 has identified key threats in the area of child sexual exploitation: peer-to-peer (P2P) networks and anonymized access like Dark Web networks; live-streaming of child sexual abuse and, minorly, Child Sexual Abuse Material (CSEM) online.

Based on a survey of LEAs and EU structures representing more than 12 000 law enforcement officials, the Operational Training Needs Analysis report “Child Sexual Exploitation” of the European Union Agency for Law Enforcement Training prioritizes the most relevant main training topics for law enforcement officials:

- Victim identification (72%)
- Combating online violence, distant child abuse, and live streaming (66%)
- OSINT and social media analysis (55%)
- Darknet (52%)
- Prevention and education (50%).

Open Source INTElligence (OSINT) is intelligence collected from publicly available sources, including the Internet, newspapers, radio, television, government reports, and professional and academic literature. Local and National Law Enforcement Authorities (LEAs), intelligence agencies, and the military commonly take advantage of OSINT. Successful operations by LEAs based on OSINT collected from Surface Web sources have forced some criminals to migrate to the Dark Web. The Dark Web is a subset of the Deep Web and includes content that is intentionally hidden and inaccessible through typical Web browsers. The Dark Web consists of several darknets including small peer-to-peer networks, as well as large, popular networks, such as I2P, Tor, and Freenet. Anonymity provides the Dark Web as an ideal environment for transferring information, goods, and services with potentially illegal intentions, therefore, LEAs are very interested in gathering OSINT on the Dark Web. However, OSINT differs quite significantly depending on whether the intelligence is performed on the Surface Web or the Dark Web. The actor conducting the reconnaissance also determines how well the reconnaissance prepares for the operation. The legal basis is very different when OSINT is carried out by a private academic researcher compared to the investigations by LE officials. Also, the line between espionage
and OSINT is thin, and caution and double-checking are advised before combining OSINT with big data analytics. LEAs must always ensure that their use of OSINT and big data analytics falls within national and international legal frameworks, including General Data Protection Regulation (GDPR) and the Law Enforcement Directive, which focus on privacy and data protection.

Child sexual abuse and child exploitation online as sexual violence including Child (Sexual) Abuse Materials (CAM/CSAM) are global phenomena. In a survey of CSAM users, 70 percent reported that they had been under 18 years and almost 40 percent under 13 years when they had seen CSAM in their lives. Many CSAM users are just not only CSAM viewers rather they also can be afraid themselves of their thoughts or feelings to make some direct contact with a real child on different online platforms after their CSAM use. Also, contact-making online or on Dark Web with other CSAM users can cause more problems by accepting their own and others’ behaviour to use CSAM even if all users do not want to tell to know one of their users. In the investigation of CAM, cross-border authority activities play a very important role. The material may be produced in one country, the viewer is from another country and the victim and perpetrator are from a third country. In addition, the investigating authority may be from a fourth country. Therefore, good information exchange and the use of Europol and Interpol are particularly important. A few years ago, a CAM database was introduced in Finland, which improves the examination of CAM and facilitates cooperation with other countries. The investigation of CAM is quite challenging as the investigation may only be carried out by an authority specifically designated for this task. Also, material containing CAM is not allowed to be recorded, so gathering evidence and presenting it in court is very challenging. The investigation of CAM can be performed on the Dark Web, of course, so that the researcher does not go to the site containing the CAM himself. However, in this situation, you must be aware of the site or link through which you can access the site containing the prohibited material. In such a situation, an OSINT could be a good way to investigate. In this way, the intelligence would be targeted at a user using a banned site or link, and the investigator himself would not have to see the CAM.

Methods

The research has been carried out utilizing theoretical research that does not immediately observe the research objects but seeks to perceive conceptual models, explanations, and structures based on previous research results and literature. The key output of the work is an explanation of how OSINT can be automated and what challenges automated intelligence possibly entails. The study also considers the challenges that arise with investigations of CAM on the Dark Web.
Child Sexual Abuse and Child Abuse Materials

Child sexual exploitation is one of the EU’s priorities in the fight against serious and organized crime. In general, cybercriminals are expected to quickly and agilely start leveraging new technologies, tailoring their attacks with new methods, and collaborating with each other in new ways. This requires LEAs to have capabilities and possibilities for how to fight cybercrime.\(^\text{12}\)

The need to protect children against sexual abuse online as a part of the fight against cybercrimes, including visually depicting children in all contents, was recognized at the European level already at the beginning of the century 2000. Europe’s Council Convention on Cybercrime 2001\(^\text{13}\) has have impacts to common criminal policy in which every member states should have national legislative and other measures to prevent all illegal material’s possession and spreading by efficient international co-operations and co-operations between states. Private industries’ measures as a part of the use and development of information technologies such as need for fast, reliable, and effective gathering of electronic evidence and, also, chancing has mentioned as an important intervention. Especially Lanzarote’s Convention 2007\(^\text{14}\) has emphasised a holistic response to sexual violence against children as a common challenge. The Convention was reinforced in 2018 with guidelines for respecting, protecting, and implementing children’s rights in the digital environment. One main goal has been to invite governments to review their legislation, policies, and practices to ensure that they can respond to children’s rights and give the highest priority to victim-focused material. Requirements for LEAs to protect and prevent children subjected to sexual exploitation or abuse based on possibilities to monitor how child sexual abuse or child depicted materials are hosted, how they can identify and locate. The recommendation to member states to actively cooperate with the Internet Corporation for Assigned Names and Numbers (ICANN) has forced to ensure that web addresses that clearly advertise or promote child sexual abuse material or any other offenses against children are identified and removed or not allowed to be registered. One important tool for large data at the national and international level is materials databases including “hashes.” LEAs must have connections to the INTERPOL database that national and international cooperation between member states would be more effective to find perpetrators or those who represents illegal materials, but also to find and collect electronic evidence from a criminal offense for pre-trial, prosecuting and trial.\(^\text{15}\)

Europol recommended in 2017 that the private sectors and platform providers should find the most effective ways to prevent, report and eliminate crimes against children in the online environment through their own services, taking into account the technological expansion in new communication channels, the growing Internet coverage and the widespread availability of mobile devices.\(^\text{16}\) The Council of Europe mentioned in 2018 that business enterprises should effectively coordinate their activities with LEA assistance, such as technical support and equipment to identify perpetrators and gather evidence using technologies available for criminal proceedings, or to facilitate child sexual exploitation and abuse material found on local servers on their platforms.\(^\text{15}\)
Many challenges exist in dealing with CSA and CAM problems with significant technological players. Coordinating efforts of many parties and actors requires also advanced technological solutions to identify, report and remove the vast existence of CSAM content online. Some players have reunited to developing systems to locate the contents of CAM by applying publicly available information, announcements, and warnings to commit a criminal offense or to help site visitors to find potential sources to get help if they were worried about their behavior. Former legislations and authorities’ jurisdictions have placed requirements on different actors to use their possibilities by side authorities work in societies. Combining earlier and future approaches together, a good possibility exists for handling the phenomenon better at the European level considering also different actors’ possible interventions widely in every stage of every society.

Recent studies recognize offenders’ manipulative ways of finding child victims, and it is important to take also consider these children’s self-reported information about various situations online. Possible harmful situations also require awareness of possible new developing ways to these contact makings and how different actors can follow and prevent these situations even if different obstacles for getting better results to prevent, detect or solve cases are quite well known. New legislation and wide cooperation at all societal levels are needed for offering better safety in the digital environment and also for children. For CSAM victims, one of the traumatic experiences is the ongoing fear of repeating the circulation of online material. The material can be used by viewed many times, which also means that the child is the victim of CSAM criminality once again. Technological issues are both possibilities that the offenders can apply, but technology also provides possibilities to protect children online. The development should also consider possible forensic tools and these technological solutions, which could help LEAs’ workload to clarify possible crimes effectively. Further research is needed to focus on almost every aspect of this phenomenon considering time preference; for example, possible new technological developmental issues, by following possible changes in victims and offenders’ behaviour or different interventions impacts to phenomenon would be important to examine.

**OSINT on the Dark Web**

When starting an OSINT operation on the Dark Web, you should carefully choose the object of inquiry, whether you want to collect information about a specific target person or one of the site’s activities and communications. Before OSINT, it is possible to take advantage of traditional means of intelligence, such as human intelligence. Once the target is selected and the OSINT begins, it would be a good idea to start the reconnaissance operation with clean equipment, especially if the reconnaissance is conducted on Tor. OSINT conducted on Tor must consider that sites are available in several languages. It would be a good idea to use several tools that specialize in gathering a certain type of information. Such tools could be, for example, Maltego and Recon-Ng. These tools
collect a huge amount of data, so time and resources must be set aside for data analysis. After analysing the information, it is possible to continue the reconnaissance operation with other forms of reconnaissance and to continue and refine the OSINT reconnaissance.

**OSINT Tools and Automation**

Gathering data from darknets is not simple and many organizations fail at that. It is therefore very challenging to automate OSINT. When automating, it should be possible to delimit very precisely the information that is desired. Opensource intelligence gathers a large amount of information from a variety of sources, so there is a risk of an excessive amount of information. Competent staff would be needed to analyse this information.

The scope of data collection is large and wide-ranging. The speed, volume, and versatility are so great that OSINT could create a Big Data problem. Tools that handle data processing tools such as Maltego and Recon-ng are becoming increasingly popular and common. However, these approaches still require regulations and a certain amount of expertise. These required settings also include a certain number of functions that cannot be automated or at least would be difficult. Retrieving data and limiting false positives to some extent can be automated.

However, it is possible to automate open-source inquiries, at least to some extent. Intelligence automation is best achieved in terms of data collection.

Automation is more challenging for analysis and automation may not be suitable for conclusions and follow-up. There are several different tools for open-source inquiry. Naini has listed on its website the eight most-used OSINT tools. The tools include Shodan, Spyse, Google Dorks, Maltego, The Harvester, Recon-ng, SpiderFoot, and Creepy.

There are several different tools for darknet intelligence for slightly different purposes. Some of the tools are darknet search engines. The Github site lists a variety of tools for darknet intelligence. These include Katana, DarkSearch, and Ahmia Search Engine. Some tools, on the other hand, acquire onion links. These include Hunchly and Tor66 Fresh Onions. Other tools, in turn, scan these same links. These include Onionscan and Onion-nmap. Some, on the other hand, index data from the Dark Web. TorBot and OnionIngestor are such tools.

Darknet intelligence is quite largely manual work. Of course, intelligence can be done in darknets, and there are good tools for this. The actual automatic intelligence cannot be performed on the Dark Web. Of course, the inquiry can be made using, for example, darknet search engines, but in that case, the intelligence should be quite well targeted and the target to be inquired about should be known. It is known that no tool could be tasked with searching for, for example, all arms trade, drugs, or CAM sites. The reason for this is clear. Creating such a fully automated tool would be particularly challenging, as the Dark Web operates in multiple languages. In addition, no precise tools are available to search for individuals on the Dark Web. Therefore, data retrieval and analysis would be very challenging to perform automatically.
The popularity of Tor has been steady in recent years. Tor is a primary target for security services in a hurry to identify and exploit vulnerabilities in browsers. For this reason, maintaining safety while using Tor is now more important than ever.24

CAM Investigation on the Dark Web

CAM investigation is a challenging and tightly regulated activity. It is also tough for investigators because its violent nature against children causes a stressful mental workload for many crime investigators. Also, the possibility of the suspect building his own defence in pre-trial and later on trial is guaranteed for example by hearings and possibilities to get information of every material which can be part of the accusation. Uncertain evidence can be agreed upon in advance by the defendant and only relevant evidence can be persuasive in the process where every facet of every issue is considered by its admissibility and weight. An investigator who gathers OSINT material must have good discretion of to which intelligence has been addressed and where, when, and how materials have been created. For investigators, it is clear to follow statutory regulations and be aware of the material’s purpose and how easily materials for example might be altered.25 In the investigation of CAM, cross-border authority activities play a very important role. The material may be produced in one country, the viewer is from another country and the victim and perpetrator are from a third country. In addition, the investigating authority may be from a fourth country. Therefore, good information exchange and the use of Europol and Interpol are particularly important. A few years ago, a CAM database was introduced in Finland, which improves the examination of CAM and facilitates cooperation with other countries.

The private sector would be well placed to assist the authorities. The Global Organization for Security and Intelligence (IOSI) could be a good help in OSINT. IOSI describes itself on its website as follows.26 “IOSI is a society-focused organization that shapes security and intelligence and is committed to promoting and improving international security. IOSI acts partly as a consultant, partly as a laboratory, and partly as an incubator. Practical solutions are being developed for existing and new security threats. These will help the public and private sectors, as well as civil society, to effectively promote public security, democracy, and human rights.” The IOSI project and its members use OSINT to gain access to information on child sexual abuse, which can help law enforcement authorities identify perpetrators of abuse and locate victims. IOSI has connections with individual OSINT experts around the world. The use of OSINT in cases of child sexual abuse can increase and help increase the number of victims found and rescued and reduce the time it takes. The challenge is again that IOSI OSINT experts are not allowed to process CAM. On the other hand, their knowledge of OSINT can be used in the activities of the authorities.

CAM investigations can be performed on the Dark Web, of course, so that the investigator himself/herself does not go to the site containing CAM. However, in this situation, you must be aware of the site or link through which you can
access the site containing the prohibited material. In such a situation, an OSINT could be a good way to investigate. In this way, the inquiry would be targeted at a user using a banned site or link, and the investigator himself would not have to see the CAM. This is how the U.S. Department of Homeland Security (DHS), among others, works. Figure 1 demonstrates how targeting an OSINT conducted by an authority to a link containing CAM could reveal a criminal viewing the CAM.

![Figure 1: Targeting OSINT to a Link to CAM Material.](image)

The U.S. Department of Homeland Security identifies the Dark Web users in the U.S. after downloading files through file-sharing services. The DHS obtained the IP addresses of several suspects who visited child pornography sites hosted on the Tor network. The researchers tracked all users who used the links to obtain an archive that contained CAM maintained on the Dark Web.

**Discussion**

Based on this study, automation of OSINT is possible, at least up to a certain point. Despite automation, the input of intelligence personnel cannot be overlooked, as automated OSINT produces a large amount of data that needs to be analysed. The analysis requires personnel who can and know how to combine the obtained intelligence with reliable and verified data. This reliable and verified data is very important, especially when the subject of the inquiry is CAM sites.

Accuracy is required if an OSINT is conducted on a known or suspected CAM site or a link that leads to such a site. It is therefore possible to automate OSINT. Automated reconnaissance is a great help when conducting reconnaissance with CAM. Such automation reduces the risk of the researcher having to see or record CAM. Automated intelligence also reduces the burden on the forensic
investigator. At the same time, automated OSINT increases the risk of criminals being caught, especially when OSINT is combined with other traditional intelligence methods.

What consequences could people’s privacy have if government-level OSINT tools were available to everyone? What risks could there be if criminals had access to increasingly effective tools?

On the other hand, in the civilian world, also companies are using OSINT to secure their commercial interests already now. Such exploitation of OSINT is a positive phenomenon and a direction of development. OSINT is, therefore, a very common and used method of inquiry. The importance of OSINT has only increased in recent years as more and more people’s lives and activities have moved online. The importance of OSINT has also increased as an increasing proportion of crime has moved online.

Automation is a significant part of today’s OSINT. However, without precise parameters, automated tools may increase the amount of data to be analysed. However, automation tools are very necessary and even essential, as it is practically impossible to perform online OSINT manually, especially very efficiently. There are challenges in automation. Not only because of the large amount of data collected but also because effective intelligence on the Dark Web is challenging. This challenge consists of the structure and purpose of the Dark Web. On the Dark Web, there are a very large number of sites implemented in different languages. Targeting your query is therefore quite challenging, you cannot find almost all the necessary data with a single search term in English.

Artificial intelligence (AI) and machine learning (ML) are likely to play an increasingly important role in OSINT intelligence in the future. If AI is the future of OSINT, how can machine vision, ML, natural language processing (NLP), autonomous machines, and robotics help the development of OSINT? These questions are topical as AI can be a perfect ally in enhancing OSINT processes when it comes to cybersecurity, military purposes, home, or even health. AI is also suitable for reconnaissance, data collection, analysis, and filtering of large amounts of data. Governments and intelligence agencies are already using AI to promote social gatherings. In particular, the military forces are counting on AI to help them succeed in the fight against terrorism, data attacks, fake propaganda, and national security.28

What kind of consequences could there be for people’s privacy if the authority-level OSINT intelligence tools were available to everyone? In particular, what risks could there be if criminals have access to increasingly effective tools? On the other hand, in the civilian world, companies are also taking advantage of OSINT intelligence to guarantee their commercial interests even now. Such utilization of OSINT intelligence is a positive phenomenon and direction of development.

Conclusions

The Dark Web gives opportunities to hide users’ identities when surfing or publishing information providing an ideal environment for transferring information,
goods, and services with potentially illegal intentions. Therefore, Law Enforce-
ment Agencies (LEAs) are interested in gathering Open-Source Intelligence
(OSINT) on the Dark Web that would allow them to prosecute individuals in-
volved in criminal activities. So, LEAs need appropriate technologies allowing
them to discover darknet sites that facilitate criminal activities and identify the
users involved. This study has been carried out utilizing theoretical research.
First, this study presents current efforts for applying various automation tools
that can be used to utilize OSINT on the Dark Web, focusing on two automation
tools, Maltego and Recon-Ng. This is followed by a discussion of the LEAs’ per-
spective on OSINT on the Dark Web with special attention to the challenges they
face when investing in Child Abuse Material (CAM) related criminal offenses.
The key output of the study is an explanation of how OSINT can be automated
and what challenges automated intelligence possibly entails. In addition, it is
considered whether automated intelligence should be performed with only one
tool or whether it would be wise to use more than one tool to get a more accu-
rate intelligence picture. In addition, the output is a reflection on the challenges
that arise in the study of CAM on the Dark Web. The study also describes the
challenges that arise in the study of CAM on the Dark Web. The biggest chal-
 lenges are related to data storage and the criminal investigation itself. CAM may
not be recorded or examined except by an LEA officer specifically designated
and trained for that purpose. The study reviews how it could be conducted using
OSINT without exposing researchers to the content of the CAM itself. The
method could be to focus the inquiry on already known links and sites. This has
challenges, but such an inquiry could be carried out by a bigger number of LEAs,
and the storage of such data would not be illegal.

References
crime-areas-and-statistics/empact.
3 Michael Glassman and Min Ju Kang, “Intelligence in the internet age: the emergence
4 George Kalpakis, Theodora Tsikrika, Neil Cunningham, Christos Illiou, Stefanos Vro-
chidis, Jonathan Middleton, and Ioannis Kompatsiaris, “OSINT and the Dark Web,”
In: Akhgar, B., Bayerl, P., Sampson, F. (eds) Open Source Intelligence Investigation
(Springer, Cham, 2017), 111-132, https://doi.org/10.1007/978-3-319-47671-1_8.
6 Bassam Zantout and Ramzi Haraty, “I2P data communication system,” In: Proceed-
ings of the Tenth International Conference on Networks, Sint Marteen, Netherlands


**About the Authors**

**Jyri Rajamäki** is a Principal Lecturer in Information Technology at Laurea University of Applied Sciences and Adjunct Professor of Critical Infrastructure Protection and Cyber Security at the University of Jyväskylä, Finland. He holds D.Sc. degrees in electrical and communications engineering from Helsinki University of Technology, and a PhD in mathematical information technology from the University of Jyväskylä.

**Iiro Lahti**, BBA, graduated from Laurea University of Applied Sciences in autumn 2021 with a degree program in information technology.

**Johanna Parviainen**, a student of Security Management at Laurea University, works as a Detective Inspector at the local police. She has years of experience in leading different crime investigations, including crimes against children. In the past three years, she has worked with economic crimes. Her Master’s Thesis explores sexual abuse of children online and CAM investigations in Finland.