Open Source Intelligence Methodology

Robert André Furuhaug

A minor thesis submitted in part fulfilment of the degree of M.Sc. in Forensic Computing and Cyber Crime Investigation.

School of Computer Science and Informatics

University College Dublin

15 May 2019

Abstract

In the role of policing, access to information is crucial, whether there is an investigation of an offence or intelligence-led investigation to prevent crime. More and more traces are left online with the increased use of digital devices and a lot of people's social life is online. More and more of this information is widely open on the internet and could be retrieved for investigative use, but to which extent is the police capable to use the potential in open source information from the internet?

When police officers understand the potential of online open source information and get the knowledge to find it, how will they do this? Are ordinary investigative principles adaptable for retrieving information and evidence from the internet?

The knowledge of tools and search operators in not enough to perform open source intelligence in most cases. There is a need for a strategy, planning and preparations before starting to gather information that takes into account a chain of custody as well as operational security and validation of the results. Furthermore, analysis of the gathered information is necessary to get intelligence out of the information.

The aim of dissertation is to describe a methodology for Open Source Intelligence on the internet that fulfil the demand for the police based on established principles and recommendations. The main approach is to review existing knowledge about principles, standards and methods from intelligence in general and investigation of digital evidence to see how existing models cover the requirements in a process for police use.

The result is a methodology presented in a model that will describe the open source intelligence process from a principal's view and from the executive's view.

Table of Contents

OPEN SOURCE INTELLIGENCE METHODOLOGY I

Robert André Furuhaug i

ABSTRACT II

1 INTRODUCTION 1

1.1 Dissertation subject 2
1.2 Method 3
1.3 Structure of the dissertation 4
1.4 Limitations 4

2 INTELLIGENCE 5

2.1 What is Intelligence 5

3 OPEN SOURCE INTELLIGENCE 8

3.1 What are Open sources 11
3.1.1 Challenges with information 15
3.1.2 Open Sources – legal issues 15
3.2 From information to intelligence 17
3.3 Procedures for Open Source Intelligence 19

4 METHODOLOGY 21

5 PRESENTATION OF A METHODOLOGY FOR OSINT 23

5.1 Requirements 23
5.1.1 Hypotheses 24
5.1.2 Requirements 25
5.1.3 Mapping 26
5.1.4 Summary 27
5.2 Strategy and Planning 28
5.2.1 Strategy 28
5.2.2 Planning 29
5.2.3 Preparations 35
5.2.4 Chain of Custody 41
5.3 Collection (Search, retrieval, and validation) 42
5.4 Processing 45

5.4.1 Validation 46
5.5 Analysis 49
5.6 Distribution and evaluation 52
5.6.1 Documentation and keeping an audit trail 52
5.6.2 Report 52
5.6.3 Distribution 55
5.6.4 Evaluation 55

6 A MODEL FOR OPEN SOURCE INTELLIGENCE METHODOLOGY 56

6.1 Comparing models 59

7 CHALLENGES 62

8 SUMMARY 63

REFERENCES 65

1 Introduction

The role of the police is to prevent and investigate crime as well as maintaining law and order to protect citizens. In all police work, information is a key issue to solve the mission given by the society. Access to information for the police coupled with the ability to utilise, interpret and adapt the information gathered has and will vary with time. At present, we are living in the information age. People leave more and more traces online with their use of digital devices like mobile phones and computers. We are social in many medias, we have accounts at many providers and when we have a question about something, we just search the Internet. Many providers of various services on the internet record a lot of data about us. Big Data analysis can be used increasingly to identify who we are and what preferences we have. This puts our privacy at risk, but simultaneously provides increased opportunities for using digital evidence in the investigation of criminal matters and to support the police with increased knowledge of people, organisations and criminal trends.

Open Source Intelligence (OSINT) provide the police critical capacity to complement and improve intelligence. If the capability to collect and analyse information from open sources is improved, it will give a great advantage in both regular investigation and intelligence-led investigations. The technical development is changing the way criminals act and we have gained a boundless and transnational development of crime in many areas such as child exploitation, human trafficking, drug trafficking, etc. It is of great importance that the police maximize the potential inherent in Open Source Intelligence and seeking new and innovative ways for preventing crime. It is also particularly important that everyone from practitioner to managers and policy-makers understand what Open Source Intelligence is, what it is not and how this potential can be exploited in favour of a more secure society, better prevention and more effective investigation.

1.1 Dissertation subject

As we all leave increasingly more information about ourselves on the Internet, this will also be the case for those who commit criminal offenses. People who do not commit offences is occasionally submitting information about criminal acts on the Internet. An example is youths sharing fight videos on social media (The Intelligencer, 2017).

A lot of all the information that resides on the Internet is more or less openly available. That does not mean that all information is just a Google search away, but that someone with the knowledge and the right tools can gain access to it. As a colleague who works extensively with information retrieval from the Internet said, "The information is out there, it is just a matter of finding it".

As there is so much information on the Internet, it will be a key issue for the police to gain the resources and the capacity to find this information. An increasingly important part of the police in the future will be to seek, find, gather and analyse information from the Internet. This applies regardless of whether it is to investigate a criminal case or to obtain intelligence to support decision making in relation to the prevention of criminal. The police have to build this capacity to meet the challenges of the future.

The military has for a long time undertaken widespread information retrieval to expose threats to national interests. Police have also to a large extent gathered information, but focus has rarely been on taking advantage of the benefits of information on the Internet. Police investigators have to a large extent tended to avoid cases or work with a technical emphasis (Nhan & Huey, 2012). In the future, open source information must become a larger part of the sources of information that the police use and this will require better methods in order to conduct this kind of work efficiently.

«Open Source Intelligence» is a term that has become established over time. The origin of the term is unclear but to gather information from open sources is not a new concept, it has just become more easily accessible with the Internet.

Over many years, Military Intelligence, investigative journalists and researchers have been performing what today is called Open Source Intelligence. Many academic institutions as well

as private and governmental organisations have performed Open Source Intelligence for a long time and have developed methods and tools for this. There are many different ways of gathering information from open sources, and how it is done will in many ways depend on which party is conducting the gathering.

Like other investigations it is important to have a methodical approach and a process in which the procedure is sufficiently documented, the evidence sufficiently analysed and the principles of Chain of Custody will be complied.

A methodology is the sum of the methods, techniques, and tools used within a discipline. A method is a systematic approach used to test something or resolve a task. A tool is an aid or technique that supports solving a task. A methodology for Open Source Intelligence will be a comprehensive approach for OSINT that can adapt to every, or at least most, investigations. The methodology must be common and cover the many different missions within Open Source Intelligence where methods, tools and techniques used may vary from case to case, but where the major methodology nonetheless is the same.

The subject of this dissertation is to describe a functional methodology for Open Source Intelligence on the Internet that fulfil the demand for the police based on established principles and recommendations.

1.2 Method

This dissertation is a theoretical assignment which explores the principles, standards and methods that are in use today and see how they are presented in different literature and in practice. The aim is to review established knowledge in order to describe a methodology that covers relevant methods chronologically through different phases and identify overlapping or repetitive processes in the work of Open Source Intelligence. The dissertation will only to a small degree address the tools and techniques that will be natural for the police to include in Open Source Intelligence as they will vary with preferences, accessibility and mission and secondly, they will change with time.

In this dissertation, references and citations are in accordance with the APA 6th standard, which is a common and widely used standard for academic papers.

1.3 Structure of the dissertation

The main part of the dissertation deals with different processes or methods that will be part of Open Source Intelligence. In order to see this in a police context we need to some degree explore the reasons why the police need to gather information from open sources. The police need updated and verified information to most of their duties, be it the enforcement of law and order, the protection of citizens, the investigation of criminal acts or by preventing crime. The same applies to other governmental institutions that perform enforcement on behalf of the society, e.g. border control, tax collection, social security benefits, etc.

In an assignment like this, which mainly concerns the collection of what will be defined as personal information, it is also important to discuss the legal aspects. Since this is not a legal assignment, legal issues will be discussed superficially as a legal discussion of the collection of personal information could be the topic of a dissertation in itself.

1.4 Limitations

The dissertation is limited to Open Source Intelligence conducted within the police. Furthermore, the dissertation will not discuss private operators' use of the Open Source Intelligence, like journalists, researchers, private companies etc. In spite of limitations as presented, the assignment is largely transferable to a variety of Open Source Intelligence in many organisations.

2 Intelligence

As the acronym OSINT points out, it is linked to the term "intelligence". Intelligence can be seen as information set in system and it is as relevant in the investigation of an offence that in projects to prevent crime. It is therefore useful to start by looking at the concept of "intelligence".

In an investigation, three distinct types of incoming data will be collected: information, intelligence and evidence (Bryant & Kennedy 2014). All three forms of data will be important for the investigation, but they have different nature and content, and they will be used in different ways.

2.1 What is Intelligence

Data is "raw information" without context. When data is processed and put in a context, it becomes information. Information from multiple sources provides the basis for analysis that is assessed against existing knowledge and thus transformed into intelligence (POD, 2014) Evidence is material suitable as a basis for prosecution and presented to a Court of Law. Both information and intelligence can be used as evidence in a Court of Law if it abides the rules of evidence.

The term “Intelligence” has different meanings in different contexts. It is used to describe a process, to describe a product of a process and to describe the organisation who conduct the process. The goal of intelligence is to create informational products to support strategic, operational and tactical decisions.

In the Norwegian Police intelligence is defined as follows: «*Intelligence is a governed process, consisting of systematic collection, analysis and assessment of information on persons, groups and phenomena to form the basis for decisions*» (POD, 2014, p. 18).

In academia, intelligence is defined mainly within two categories. The first category is where intelligence is described as a process, defined as "*the systematic and purposeful acquisition, sorting, retrieval, analysis, interpretation and protection of information*" (Harfield & Harfield, 2008, as referred in Staniforth, 2016, P. 23). This form of definition is in line with the definition of the Norwegian police.

The second category describes intelligence in relation to information. Intelligence is described as «*information derived from many sources that has been recorded, graded and evaluated, in short, intelligence is information with meaning*» (Bryant & Kennedy, 2014, p. 124) and as «*information that has been given some added value after being collated and assessed*» (Kleiven, 2005, p. 40)

In the intelligence doctrine of the Norwegian Police, some requirements have been described to make good intelligence (POD, 2014, p. 19-20).
Objectivity and integrity: Intelligence must be open-minded and objective. This requires integrity that is necessary for the quality and credibility of the intelligence. Intelligence shall usually rely on multiple sources. Uncertainty in reviews shall appear clearly.

Timely. Intelligence must be delivered in time to have relevance for the principal's decisions. This includes that time of delivery of intelligence may be governing even though a broader spectrum of sources and a further processing could increase the quality on the intelligence.

Centralised control is necessary to both reinforce priorities and ensure efficient use of resources.

Systematic and considered application of gathering methods based on good knowledge of their possibilities and limitations is a prerequisite for a retrieving relevant data and information.

Documentation and traceability are important for preserving an audit trail of both retrieved information and reviews, to prevent unnecessary reporting, to avoid manipulation attempts, and to safeguard that intelligence products and information becomes part of the organisations' knowledge and not depending on the individual officer.

Protection of sources and shielding. Sources must be protected from risk and manipulation. Some retrieval methods necessitate the need to shield the identity or capacity of the source. This is safeguarded through sanitisation, where information that can lead back to the source is removed or rewritten.

Availability and dissemination. Intelligence products are to be made available to the client. In addition, intelligence should be shared with relevant internal and external stakeholder as far as it is security-justifiable.

Prioritization and adaptability are a condition of the relevance of the intelligence when situations evolve, and decision-makers need change. Intelligence needs, products and methods must be considered on an ongoing basis. This requires the ability and willingness to see weaknesses and learn from experience.

With the definitions that are presented, intelligence can both be something you do and something you have. Regardless of whether you see intelligence as a process or based on information, the key issue of intelligence is information. Whether the information is to be used to prevent crime, prevent threats against the nation or the investigation of committed offences have no effect on the definition of choice. A closed definition which the Norwegian police uses does not point in the direction that this also deals with the collection of data to be used as evidence in court. It is therefore preferable to use a more open definition that relates to information as product and does not provide guidance on what the information will be used for.

3 Open Source Intelligence

Open Source Intelligence is a demand-based method of collecting information from open sources. Open sources can at the outset be all sources that are widely available. That a source is widely available means in principle that everyone should have lawful access to the information. That is, information that has been accumulated in case of illegal access (hacking, etc.) or information which is limited to certain persons in certain organisations falls outside the term.

The term Open Source Intelligence must be delineated against the simple Google search and Internet lookup police do as part of the job. Finding a phone number on Yellow Pages is not Open Source Intelligence although it is on Internet and it is an open source. Open Source Intelligence cannot be considered a separate subject, neither within the police nor other professions, but a method of collecting information regardless of what its purpose is. To the extent that the investigation is a subject, OSINT will be a method to obtain information when working with the subject. Intelligence is a process of gathering and analysing information, and Open Source Intelligence is a method for gathering and analysing information from open sources.

The purpose of collecting information from open sources is first and foremost to increase the total intelligence with available, relevant and precise information. The benefits of OSINT in relation to many other forms of information retrieval are many. Some of the most central advantages are according to Hassan and Hijazi (Hassan & Hijazi, 2018, pp. 15-16):

> Less risky: Use of publicly available information to collect intelligence is of low risk compared to other forms of intelligence such as the use of informants, undercover agents and other forms of physical presence in a criminal environment.
>
> Cost effective: Collecting information from open sources is generally less expensive than collecting information from other sources. OSINT will be able to provide more intelligence for the money. It is important to point out that OSINT does not necessarily

provide the same information as other techniques, but it can still provide the information that is needed.
Ease of accessibility: Open sources of information are always available regardless of location. Access to Internet based open sources only requires a computer with an Internet connection and sufficient knowledge to find what one is looking for.

Aiding financial investigators: OSINT allows specialised governments to find tax evaders and social security scammers etc. Information from social media etc. can reveal signs of spending that are in excess of income, for example.

Fighting against online counterfeiting: OSINT can be used to detect counterfeit products and fraudulent services and direct the police to close Web pages and uncover perpetrators.

Maintaining national security and political stability: Threats from terrorists and others can be uncovered with the help of OSINT and groups with radical attitudes can be mapped and monitored.

This is also supported by Ramwell, Day & Gibson (2016) who describe how the Metropolitan Police already in 1999 pointed out the benefits of Open Source Intelligence. They noted that the police officers quickly saw that speed, efficiency, availability and costs were great benefits OSINT.

Open Source Intelligence can also be an important support tool for various governmental agencies investigating different categories of offences. For the police, it can be used in the investigation of most types of offences as well as being an important tool for those parts of the police and the armed forces working for national security and political stability. Furthermore, the tax administration, Social Security Agency, and other governmental agencies may also use Open Source Intelligence as a tool to gather intelligence on targeted actors. Private companies that provide various services will also be able to use Open Source Intelligence. It can be banks, internet and phone companies, oil companies, etc. that are experiencing various threats to their assets or their infrastructure.

Using open sources is not a new phenomenon. In particular, the military has led the way in implementing Open Source Intelligence as part of information gathering. Since the 1940´s, intelligence has relied on open sources to gather information (Bradbury, 2011, as referred in Ramwell et.al., 2016). I Pre-Internet time, Open Source Intelligence consisted of collecting information from newspapers, news broadcasts, etc. from abroad and was particularly applicable to military intelligence. During the Cold war, both sides of the conflict built up large repositories of newspapers, magazines, books (Schaurer & Störger, 2013).

Books, journals newspapers, fliers, articles, maps, documents etc. have always been a source of knowledge for those who conduct research in some form. The game-changer was the introduction of the Internet because access to the open information become much easier. At introduction, the internet was probably not a place for Open Source Intelligence, but as it grew, it became a significant factor. In NATO's OSINT Reader From 2002, they estimate the size of the Internet to Six Terabytes (NATO, 2002). Today, the index of Googles alone is well more than 100 000 000 Gigabytes (100 000 TB)[1]. The year 2002 is early in the Internet history from when it began to spread with World Wide Web in 1991 to present time even back then, the internet was a major factor for Open Source Intelligence already then. Today, a large amount of available information is still presented in printed form, but more and more of the printed text also exists digitally, such as newspaper articles.

The next game-changer in OSINT came with the use of social media. To begin with, this was a trend especially teenagers used, but today everyone from grandma to grandchildren are present on different social media platforms. The use of information from social media is especially valuable to those working against specific individuals and specific groups of people. Admittedly, not everybody is present on Social media, but social media has become an essential part of life to an essential part of the world's population. Even those who do not have their own social media account may be referenced by others, such as children, friends etc, thereby revealing insight into their lives. Facebook alone had pr. 2. quarter 2018 2.23 billion registered monthly active user accounts[2]. This represents nearly a third of the world's population. Taking into consideration that one must be thirteen years old to register for a user account (although many youngsters under thirteen years register anyway) it is not far off that

[1] https://www.google.com/search/howsearchworks/crawling-indexing/

[2] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

half of the world's population who fills the requirement to have a user account, have a Facebook account.

Even many of those who choose not to have a user account on Facebook will be registered somewhere on the internet. It can be information they've posted themselves or that others have posted. A results list from an event can provide a name, searches for the same name with the organisation that arranged the event may provide an email address and a phone number, a further search on this may display a user account or a user name etc. This way one may obtain new pieces of information that fit into the puzzle.

Edmund Lockard described what has become known as Lockard's Exchange Principle, which has been an important principle for forensics. The principle, that says that «*when a person or object comes in contact with another person or object, a cross-transfer of materiel occurs*» (Saferstein, 2007, as referred in Årnes, 2018, p. 2), describes that all interaction will leave tracks. This also applies to the highest degree on the Internet, if we only interpret "material" as data. Every user of the internet will leave traces of their use and much of that information will reside open. With the huge access to, and the use of, the Internet in modern society, any criminal or offender will spend substantial time on the Internet. Even if they do not commit their offence over the Internet, they will use it for other services and this is where information can be found. It should be equally natural and equally integrated in the everyday life of every investigator to look for the information they need on the internet. In that context, OSINT comes as an essential method of collecting information from open sources on the Internet.

Those who grew up before the Internet became a factor in everyday life will have a life outside the internet, but they have also established a life on the Internet. Those born under the Internet will in return have their whole life on the Internet, placed there by parents, family, friends, etc. Basic OSINT Seeking to find and document these digital footprints accrued over time from many sources from the targeted individuals.
(Ramwell et.al., 2016).

3.1 What are Open sources

Open sources will basically be any information that has a theoretical potential to reach anyone. Such a broad definition, in many respects, will not have any practical significance, but there is no need to refine the concept out over Open versus Closed.

NATO describes four distinct categories of open information and intelligence (NATO, 2001, p. 2-3):

> **Open Source Data (OSD).**
> Data is the raw print, broadcast, oral debriefing or other form of information from a primary source. It can be a photograph, a tape recording, a commercial satellite image, or a personal letter from an individual.
>
> **Open Source Information (OSIF).**
> OSIF is comprised of data that can be put together, generally by an editorial process that provides some filtering and validation as well as presentation management. OSIF is generic information that is usually widely disseminated. Newspapers, books, broadcast, and general daily reports are part of the OSIF world.
>
> **Open Source Intelligence (OSINT).**
> OSINT is information that has been deliberately discovered, discriminated, distilled, and disseminated to a selected audience, generally the commander and their immediate staff, in order to address a specific question. OSINT, in other words, applies the proven process of intelligence to the broad diversity of open sources of information, and creates intelligence.
>
> **Validated OSINT (OSINT-V).**
> OSINT-V is information to which a very high degree of certainty can be attributed. It can be produced by an all-source intelligence professional, with access to classified intelligence sources, whether working for a nation or for a coalition staff. It can also come from an assured open source to which no question can be raised concerning its validity (images of an aircraft arriving at an airport that are broadcast over the media).

Textual sources of open information may be in printed form or digitally published. For the "open source" criterion to be valid, it must at least be information that everyone has the

potential to see in a lawful way. Furthermore, the information must be based on freely, widely available Sources (Akhgar, 2016) but "open" is not the same as "free"; «*The word 'open' in open source intelligence must not be confused with the word 'free'*» (Gibson, 2016, p. 81). There are many sources of open information behind payment services.

In NATO OSINT Reader, they refer to a definition of open information from «Director of Central Intelligence Directive»

> "*Open source information for purposes of this directive is publicly available information (i.e., any member of the public could lawfully obtain the information by request or observation), as well as other unclassified information that has limited public distribution or access. Open source information also includes any information that may be used in an unclassified context without compromising national security or intelligence sources and methods. If the information is not publicly available, certain legal requirements relating to collection, retention, and dissemination may apply.*"

(NATO, 2002, p. 9)

According to this definition, Open Source Information is «publicly available information». The criterion to be seen as publicly available is if any member of the public could lawfully obtain the information by request or observation. In addition, it defines unclassified information that is subject to limited public distribution or access as well as information that may be used in an unclassified context without compromising national security, etc. as Open Source Information. They thereby operate with a relative wide description of Open Source Information.

Federal Bureau of Investigation (FBI) defines Open Source Intelligence as:

> *«...the intelligence discipline that pertains to intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely matter to an appropriate audience for the purpose of addressing a specific intelligence requirement*».

(Staniforth, 2016, p. 24)

NATO OSINT Reader also operates with the concept of "Grey literature»:

«*Grey literature, regardless of media, can include, but is not limited to, research reports, technical reports, economic reports, trip reports, working papers, discussion papers, unofficial government documents, proceedings, preprints, research reports, studies, dissertations and theses; trade literature, market surveys, and newsletters. This material cuts across scientific, political, socio-economic, and military discipline*"
(NATO, 2002, p. 9)

Here they include research reports, technical reports, financial reports, trip reports, etc. Much of what is described as grey literature could also be defined as Open Information, so there are no clear boundaries between grey literature and Open Source Information.

Another approach is to describe information from open sources by clarifying which sources are closed. A closed source of information is a source that does not have the potential to reach everyone. Closed information is only available for a limited circle of people. It may be the board of a company, employees of an organisation, friends of a person on Facebook, members of an association etc. As long as the information is not potentially available for everyone it will be a closed source. However, it is conditional on the fact that there must be an actual limitation of information.

As an Example of the distinction between open and closed information we can look at Facebook. To get any particular information from Facebook, beyond the small pieces of information we can get by searching in Google, pipl.com or similar, one has to be logged in to Facebook. Facebook is basically open to anyone who fulfils the requirements for creating an account. The information anyone can find by searching inside the Facebook, can be considered as open information. It is freely available, public and access to the information is legal. The information you can access only by following a page, as a member of a group or friend with a person will initially be considered closed information. There are some exceptions in terms of pages and groups, and that regards to pages that everyone freely can choose to follow and groups that anyone can become a member of without any form of approval or screening. Information from such sites and groups is also freely available to everyone.

3.1.1 Challenges with information

One of the challenges in terms of information is that it can both be constant and volatile. Information is rarely etched in stone, but it can still be so widely distributed that it becomes permanent. Information in a single book will be lasting as long as the book exists. If the library holding the book (the only copy in existence) was to burn down, then the book will be gone and thus also the information contained in it. Today, a lot of Information is distributed in such a scale that it will be secured against such circumstances. If all non-closed information is considered open information, then what we can freely observe is also considered as open information. A lookup on a single lamp post is an open source since it is available to everyone, even though only a few people realistically will see it.

The problem of ephemeral information is precisely that it is volatile. How can such information be validated by others and not only will be left as an assertion? When scientists write their articles based on different sources, these are sources that will be available to others. Anybody can verify the sources and validate whether the information is correct. However, volatile information cannot be validated when it is absent. The existence of the information must be documented. That is a form of preserving the information. What we observe can be preserved with images, with the limitation that it is only what can be read out of the image that is currently open information. No one else can verify that it really was that way. The same goes for information we find in the open forums on the Internet, on social media, and other places where users can publish and delete information. If we find open information in such places that we wish to use further in some context, we have to document that it has been there. It may have been deleted when someone is entering the same site to validate that what we have found is true.

3.1.2 Open Sources – legal issues

The term open sources of information are no legal size. There is no legal definition of what open or closed sources of information is. That does not mean this is an area of no legislation. Referring to NATO's definition of open sources of information, legal access is required. When open information denotes a lawful access, it means in practice that one can obtain the information without committing an offence. That does not mean that the information is provided and made open without any offence. Wikileaks is an organisation that publishes classified information they receive from various sources that have either stolen the

information (insiders), broken into computer systems (hacking) or similar. This information is to a large extent used by journalists. The same applies to Panama Papers, where large amounts of alleged stolen information from the law firm Mossack Fonseca In Panama was published on the internet[3]. These documents have largely been used by tax authorities in many countries to track assets that is suspected to be withheld from taxation. In other words, it is not the original access to information that must be legal, but the individuals´ access to the information that must be legal.

Open information is being delineated against closed information. However, it does not mean that it intends an offence to access closed sources. Although it has been clarified that access to the information must be legal for it to be considered as open information, it does not mean that any closed information is illegal. It might be other limitations to the use closed information. Much such information may be used, shared with others or published without it being an offence, but there may be violations of internal rules, terms of employment, board of Directors, etc.

A borderline can also be drawn between what is available of information and what is coming from open sources. Open sources will in most cases not be the only source of information but be part of the overall picture. This applies regardless of whether it is a journalist, an intelligence officer or an investigator who collects information. The reason that it is important to distinguish between what is open information and what is closed is that one, regardless of role, can openly refer to open information because it is available to everyone. Reference to closed sources must, however, be considered in each case.

Regardless of whether information comes from open or closed sources, the police will have a limitation in collecting information from people. If a person has a blog in which she publishes about her private life, an investigator will be able to follow her blog without any violation of the privacy laws of EHCR Art 8 and national laws derived from this provision. Data Protection Rules however, applies in the case of any storage and use of personal data by the police regardless of whether data is collected from open or closed sources (Sunde, 2018).

[3] https://www.icij.org/investigations/panama-papers/

Therefore, the police cannot collect, store, and use personal data about citizens without a legal reason for their storage and use.

Although there are no legal limits between the open and closed sources of information, the legal issues have absolutely a meaning, both for the access, storage and use of information, especially for governmental authorities, that one must be aware of.

3.2 From information to intelligence

The reasons an investigator would want to collect information from open sources are many and varied. It may be because it is the only way to obtain the information or it may be because the information has been obtained from closed sources that cannot be revealed, hence seeking to verify with information from open sources, that may be used further in the investigation and later, in Court of Law. Intelligence in government agencies is governed by a process that exists in different variants and is often called the Intelligence process or the Intelligence Circle.

The intelligence process is found as mentioned in different varieties. Norwegian police operate with a four-phase process:

- Management and prioritisation.
  It starts with the fact that decision-makers has a need for knowledge. In dialogue, the intelligence they need, which are broken down to requirements, are added to the basis of their continued work.
- Collecting.
  It includes the use of various retrieval methods and the dissemination of information for analysis and assessment. Data and information collected should answer the identified requirements.
- Analysis and Assessment.
  Information is being converted to intelligence by processing, analysis and assessment, before the intelligence products are processed.

- Dissemination.

  The intelligence product is communicated and distributed to the principal in the correct format at the appointed time. When distributing, it is important to keep the focus on information security.

(POD, 2014 p. 26)

The FBI presents a circle of intelligence that also has four phases:

1) Requirements management – identifying what we don`t know
2) Collection – gathering information on what we don't know
3) Production – answering the question, and
4) Dissemination – getting the answer out to the right people, whether it be the President of the United States or the patrolman in the streets.

(Mueller, 2004)

Rogers (2012) presents the intelligence circle as a circle with five phases:

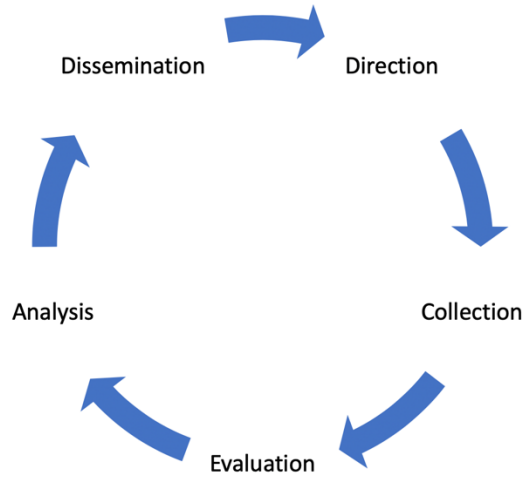

Fig. 1. The Intelligence Cycle. (Rogers, 2012, p. 132)

The intelligence circle starts with the phase "Direction", covering the identification of intelligence required, to data collection and evaluation, further on to the analysis and finally to distribution of the result. Rogers points out that distribution has been a challenge for all Law enforcement agencies due to ownership and lack of trust in others (Rogers, 2012).

As we can see, these processes have many common denominators and the content is largely the same, although the sections and the naming are different.

The transition from information to intelligence will mainly take place in the analysis phase. This is where the data collected is reviewed and put in system to provide an increased understanding. In the NATO OSINT Handbook the difference between (Open source) Information and (Open source) intelligence is that Intelligence is information that is "*deliberately discovered, discriminated, distilled, and disseminated to a selected audience, generally the commander and their immediate staff, in order to address a specific question*" (NATO, 2001, p. 2-3). In the analysis phase, the comparison and distillation of the information will transform into intelligence.

3.3 Procedures for Open Source Intelligence

Although the principles of Open Source Intelligence are fundamentally similar regardless of whether it is a journalist who use the method or if it is the police, there are also some distinctive differences. The methodology will be different to some extent but however, there are some universal points that should be included in any methodology for Open Source Intelligence. One must have a purpose, defined as a requirement. Based on its purpose, one must devise a strategy and a plan to achieve the goal. Furthermore, one must map what you know and what information is missing to achieve the goal. In a police context, documentation of everything that is done is even more important than in a journalistic context. It is inextricably linked, to the requirement of an Audit Trail and that not only the information, but also the source of the information is to be documented for use in trial. Furthermore, a stricter validation of the information is required for evidence in a criminal case. As the information, and perhaps the result of analysing the information, which are mostly central to the police, it is important to document an audit trail from the start. One will never know for sure what turns out to be key evidence and without an audit trail, good evidence can be reduced to a weak assertion.

4 Methodology

To facilitate a rigorous process, a good, structured and targeted work is necessary. This also applies to the highest degree in OSINT. A good process should ensure that the investigator does not skip steps in the investigation and that Chain of Custody is taken care of in a proper manner all the way. To ensure a good process, a methodology must be required as the basis of the process. The methodology describes the phases to be reviewed and what is included in each phase so that it may be used as a manual through the process.

A methodology is the study of methods within a discipline, whereas a method can be defined as a systematic approach to resolve an issue or task. A tool is an aid or technique that can help you complete a task. The methodology is a comprehensive approach that can contain several methods, techniques and tools (Bjerknes & Fasing, 2018).

The purpose of the methodology is to define a structured investigation to ensure that it remains forensically sound. The investigation can be regarded as forensically sound if it complies with the established principles, standards and processes of digital investigation (Flaglien, 2018). A methodology for the investigation of digital evidence must be based on principles of digital forensics, and common Law Enforcement and industry practices.

Information from open sources should be used both as a basis for ordinary investigations and intelligence-led investigations. Where the information is used as a basis for intelligence-led investigations, the information will not necessarily be used as evidence in court. However, it is important that an Audit Trail are maintained so that decision-makers can rely on the information that the decision is based on being correct. When the information is to be used as evidence to for prosecution, the authenticity and integrity of the evidence must be safeguarded in such a way that there can be no question of the value of the evidence.

A good methodology will also help us to avoid legislative challenges during the process of collecting information. Furthermore, it will keep us targeted and ensure that we adhere to the hypothesis and information requirements that underlies Open Source Intelligence. A Good methodology that based on the principles and standards of the Investigation of digital

evidence will ensure that evidence is lawfully secured and that the evidence is examined in accordance with best practices to clarify the origin and whether it is tempered with or placed to mislead the police in the wrong direction or to direct suspicion against an innocent person.

A methodology should describe the purpose of the task, the process, the different phases and their methods and tools and how the result should be presented and distributed. There are various models that presents a methodical process for working with digital evidence. It is published several standards and guidelines for the Digital Forensic Investigation Process. ISO/IEC 27037 and NIST SP 800-86 presents standards for the investigation of digital evidence (Dilijonaite, 2018). There are also published Guidelines like ACPO Guidelines IOCE Guidelines and Electronic Evidence Guide (EEC) which advises how digital evidence shall be handled. Both standards and guidelines will be advisory. That means that evidence will not automatically be rejected because a given standard or guideline has been followed, but they are designed to ensure a forensically sound investigation of digital evidence. An investigation aims to provide evidence to enable prosecution and this evidence shall be relied on in a Court of Law. Guidelines can thus be considered to describe a Chain of Custody that ensures that the evidence is preserved in a way that safeguard authenticity, integrity and reliability of the evidence. Although the standards and Guidelines is different, they don't stand against each other. It is rather so that they complement each other. One can and should be able to use several of these when working with investigation of digital evidence.

A methodology that describes a process for working with information retrieval from open sources must also build upon the principles applicable to the investigation of digital evidence. Although the information is open, preserving the data should follow the standards and guidelines for preserving digital evidence, especially when the information is to be used as evidence in court. Since a methodology for Open Source Intelligence should cover the process of acquiring information from open sources for use in both reactive and proactive investigation, inspiration and knowledge will be taken from both the intelligence process and guidelines for the investigation of digital evidence, as well as fundamental investigation principles.

5 Presentation of a methodology for OSINT

5.1 Requirements

There are many models of the intelligence circle or the information circle. Some have Direction as a start point while others have Requirements as a start point. Although the models are different, the content is in a greater sense the same. They start with a need for information that leads forward in the process. In this methodology, a model with Six phases is presented. The first phase is Requirements.

An investigation can be both reactive and proactive. The reactive investigation is performed based on a review of a (or several) criminal offence. The aim of the investigation is to find the necessary evidence to enable prosecution Necessary evidence might be to find an unknown suspect, to clarify how the crime was conducted, to clarify the motive. A proactive investigation is also called intelligence-led investigation (Gibson, 2016), in which intelligence has either provided information that a criminal offence is being planned or that a criminal trend is spreading and he police would want to influence a change of direction in order to stop this trend developing. There are some distinct differences between intelligence and evidence. The purpose of intelligence can be wide and varied and used both in investigation, rescue operations and crisis management. Evidence is solely for the purpose of assisting in court to clarify the enlightenment of the case (Sampson, 2016).

There will be a demand for information from various sources in all cases. Hypotheses will be formulated based on the information analysed by the police in an attempt to verify or disprove them. To verify or disprove a hypothesis, information is needed. An informational requirement is defined based on the hypothesis and what information is required to confirm or disprove it.

The Requirements phase in this model will describe the mission, setting hypothesis and Requirements management.

5.1.1 Hypotheses

A problem can often be formulated as a question. From the question, one or more hypotheses may be inferred. A hypothesis is a possible explanation, "*an idea or explanation for something that may be true but has not yet been completely proved*" [4].

For all investigations a hypothesis as a proposition made as a basis for reasoning without the assumption of its truth and supposition made as a starting point for further investigation of known facts (Staniforth, 2016)

The idea of hypotheses is that the focus of what needs to be examined is narrowed down and that it provides direction for future work. The development of hypotheses is important for planning, governance and management of the task. Hypotheses are a guideline for identifying requirements and sources of data. This structures the work and contributes to a common platform and understanding for everyone involved. (POD, 2014)

In an investigation, reactive or proactive, it will be necessary to develop alternative hypotheses as this offers more opportunities for clarification, increasing objectivity and highlights different ways of development. Alternative hypotheses should be substantially different from the hypothesis the work assumes, but at the same time be probable (POD, 2014). Alternative hypotheses should be alternate explanations that may be probable if the main hypothesis is not viable.

The making and use of hypotheses are a widely recognised technique among investigators who can be used to assume the most logical or likely explanation of how and why a criminal action has been committed. Likewise, it can be used to assume the most logical or probable explanation of who has performed such an action. (Staniforth, 2016)

Verifying a hypothesis is not always possible. The alternative procedure is to falsify the hypothesis. Falsification is a detection that a theory is wrong or untenable. Falsification is the opposite of verification.

[4] https://dictionary.cambridge.org/dictionary/english/hypothesis

Today, falsification is an important concept in the philosophy of science, particularly with Karl Popper, who argues that there is a momentous asymmetry between falsification and verification regarding general statements and natural laws. A complete verification of natural laws is impossible if they (simplified) have the logical structure "all A has the property E". However, it is in principle logically possible to falsify them: the assertion "All ravens are black" is in principle falsified if we find a raven that is not black. Although all ravens we have seen so far, are black, we cannot know that there has never been, or will be discovered, ravens who are not black.

For Popper and his followers, it became a methodical principle in all science to increase the degree of falsifiability. A law or generalization (theory, hypothesis) that resists attempts at falsification, thereby increases its acceptability or credibility. The more efficient we search for falsification or possibilities for falsification, the more efficiently we promote the growth in knowledge[5].

When the hypotheses are set up, they will provide direction for the further work of information gathering to verify or falsify the hypotheses. This also applies to information retrieval to be used as evidence in criminal proceedings, as the evidence can underpin or confirm a possible explanation.

5.1.2 Requirements

Based on the established hypotheses, various explanations related to these are explored. The subsequent requirement for information will aid towards verifying or falsifying these hypotheses.

Example: Police have registered repeated cases where young men have been ambushed and beaten by groups of five to ten youngsters. A hypothesis can be that there are a group of youngsters who are seeking out lone young men within a given area and they beat them up to mark themselves as a group, an entity, to highlight a territory or similar. An alternative hypothesis could be that there are two or more gangs operating within an area and who attack

[5] https://snl.no/falsifikasjon

lone gang members as and when they come across them and that this signifies a battle for territory. The police must then seek information to verify or falsify the hypothesis until they are left with the most likely or a verified hypothesis. Only when it is confirmed that one of the hypotheses is verified will it be possible to direct effective responses to the problem.

Whether the police should map a trend to prevent offences or investigate a punishable offence we can divide the required information in two parts: what we know and what we do not know. The information we do not have which can fulfil the complete picture is the needed information. Mapping and clarification of what we do not know is what the former director of the FBI, Robert Mueller calls Requirements Management (Mueller, 2004)

The requirement management can be described as follows: Clarify the information needed to verify or falsify a hypothesis, then subtract the information already acquired, and what is left is the information requirement.

To provide direction for the further phases of the process, the requirements must be consolidated and delineated. To describe requirements with "anything that can confirm (or disprove) that.... is the case" becomes to abstract. In the example above, specific requirements may be: «Mapping of persons related to gang A and gang B, » «information about any conflicts between gang A and B», «Information about conflicts between people connected with gang A and gang B» and «information about if people from gang A and B is registered as victims in assaults».

5.1.3 Mapping

A thorough survey of information will be a central part initially in an Open Source Intelligence Investigation. The first part will be, as mentioned above, to get an overview of what we already know. In larger cases this phase could include structuring and analysing the information the police already possesses, from police databases, from police records (interrogation, reports, etc.) and police Intelligence systems.

A definition or clarification of the term 'information' will not be necessary as part of the mapping, but a description of which information is required in the investigation is of great importance.

The requirements will vary from case to case and it may vary whether it applies to mapping a criminal trend or to an investigation of a case. In the case of investigation, requirements will largely be related to finding a suspect, uncover the evidence that may prove that they have performed the offence, identify motives, underlying causes and conveying circumstances etc. Retrieving information from open sources will have relevance to all parts of the investigation.

Center for Security Studies (2008) estimates that information from open sources represents between 80 and 95 % of all information used by the Intelligence Community.

5.1.4 Summary

During the requirement phase, hypotheses will be established, and information requirements mapped. This will pave the way for the next phases in the work. Once the information requirements have been defined, a strategy for gathering information from open sources is formed.

5.2 Strategy and Planning

An important phase in the Open Source Intelligence process is strategy and planning. There are many reasons why the police need to collect information from open sources. It will, in many contexts be related to the investigation of one or more criminal offences or to prevent criminal offences (Gibson, 2016). Open Source Intelligence must be governed by a purpose. Purpose is defined by the case, the hypotheses and the requirements. The purpose may also be based on a project. That could be a project based on an overall strategy that defines the priorities, nationally or locally, of the police. In a methodological context, the strategy is specific and provides the basis for the planning and preparations that will govern the work of providing the required information.

5.2.1 Strategy

Strategy closely correlates with planning and therefore can be considered together. Strategy and planning form the basis for the next phases of collection, processing and analysis. Different models include or describe the strategy during the phase «Direction» if strategy and planning are included in the model at all. Strategy shall set the direction for the future work, but it must build upon the prior work of establishing hypotheses and requirements management. The strategy will be different from case to case although the methods one uses are the same.

Open Source Intelligence will be one of several ways to provide information to a broad range of information. Strategy for Open Source Intelligence must be based on verifying or falsifying the hypotheses by covering the information requirement defined while the whole process of OSINT is building on the purpose of the investigation. Strategy will therefore describe the purpose of collecting information, what information is required, which sources are most relevant to seek information from and how the information is can be preserved, documented and validated.

Strategy must be specific in form and content so that all parties involved understand the purpose and direction of the mission.

5.2.2 Planning

Planning will connect closely to the strategy. Where the strategy provides direction for the investigation, the planning will facilitate a move in the right direction. In many situations, there will be information that changes the focus, one must search other places, or must follow other threads than one thought at first. Nevertheless, planning is a key part of Open Source Intelligence because it's not just about planning where to think about finding information, it's about preparing the entire process of collection, processing and analysing information, considering equipment, accounts, legends, operational safety, etc.

Thorough planning of how to identify the relevant information required to answer the questions and the process of finding and preserving this data is an important first step to extract information that has the necessary quality and accuracy (Gibson, 2016).

The fact that information can be found in open sources does not mean that it is easy to find or access. One must consider what information may bring the investigation forward, where that information can be found and how it can be extracted. Information will be found in many formats and how to deal with all the different formats should also be included in the planning.

As part of the planning it is important to describe the concepts of grouping information, validate information and value of information. This will be useful to bring further to the phases of collecting, processing and analysis.

5.2.2.1 Grouping of information

In a model where intelligence is derived from information, information must be able to be categorised in a systematic way. In a standardisation of data, you need to define a few regular, common data "objects" that the investigation will be associated with. These can be categorised as event objects (such as robbery, assault, etc.) and static objects (like persons, vehicles, buildings, etc.). By identifying each unique object, a complete list of all relationships between the different objects can be created.

To group information, it is necessary to consider which grouping might be appropriate. A standardisation of grouping of information will be useful for the later analysis. TechUK identify four points of research to be examined which can be applied to all investigations.

These four points of research are Person, Object, Location, and Event, abbreviated to POLE (TechUK, 2014). At least one of these must be present to have an enquiry. Intelligence obtained from any source will be able to be attributed in the terms of the POLE data model. In the UK this is in line with the guide for Authorised Professional Practice from UK College of Policing and it is debated whether this should be a common standard for the police in the UK (Ramwell et.al., 2016).

The four points of research described above in the POLE model can be used to group entities. Entities are objects such as people, organisations, places, etc. that emerge in the material. Entities can also be identities, email addresses, nicknames, IP-addresses[6] and more (Gibson, 2016). All of these entities can be grouped into people, objects, locations, and events. In other words, an entity is a person, object, place, or event described in the investigation.

People

People will often be the most comprehensive entity in most cases. There is always a human being behind a crime, whether it is fulfilled or at a planning stage. In many cases there are several persons involved in both the crime and the preparations in advance. That makes the people entity in any investigation, reactive or proactive, the most central object of investigation.

Intelligence on persons in an investigation will be both information about the individual person in the case and the various people's relationship with the other entities.

People as information objects can be both identified and unidentified. A known suspect, a witness, or a victim will be identified person entities. An unidentified person entity might be an unknown suspect or a witness who is described in the case without the knowledge of who the witness is. That person can be described as a witness in the case based on information from others, video surveillance or similar.

Objects

[6] An IP address is numerical label assigned to any device connected to IP-based network which is used for host or network interface identification and local addressing (https://en.wikipedia.org/wiki/IP_address)

Objects are entities which do not fall into the other categories, People, Location, or Events. In other words, objects can be almost anything. A vehicle, an animal, a weapon, money, etc. will be objects in such a classification. Buildings can be objects, for it is not given that a building and a location is connected. A location entity can be a location without buildings and secondly, a building can be an object entity without it being attached to a location, i.e. a building can be described in detail without identifying its location.

Objects do not have to be physical sizes. In particular, the Internet will identify many objects that do not exist physically but exists only digitally. An IP-address, an email address, a domain name, a user name, and a Facebook ID are all items that only exist digitally, but they are still important objects in the investigation.

Location

Locations are entities that can describe an area, a place, etc... The most central location will often be the scene of a criminal offence. Some criminal acts take place over several crime scenes, e.g. theft of a wallet (scene 1) that contains a credit card that is used for to withdraw money (scene 2), perhaps several times (scene 3, 4, etc.).

It can also be a phenomenon that occurs several places over a period of time, like serial robberies or serial rapes, where the nature of the case makes them seen in context.

Other locations in an investigation can be places of residence for involved persons (suspects and victims), places of planning and preparation, arrival and escape routes, observation posts, etc.

Locations related to digital evidence will also be Location entities in an investigation. There may be location of base stations on which mobile phones have been actively or passively connected to, or base stations covering a crime scene or other central locations. There may also be locations extracted from EXIF[7] data from pictures in mobile phones, etc.

[7] EXIF is short for exchangeable image File, a format that is a standard for large-size interchange information in digital photography image files using JPEG compression. Almost all new digital cameras use the EXIF annotation, greating information on the image such as shutter speed, exposure compensation, F number, what metering system was used, if a flash was used, ISO number, date and time the image was taken,

whitebalance, auxiliary Lenses that were used and resolution. Some images may even store GPS information, so you can easily see where the images were taken! (http://exifdata.com)

Events

Events are nearly as central to an investigation as people. This applies to proactive investigation when historical events can reveal something about the anticipated events one wants to prevent from happening.

In the same way that the crime scene is the most central Location entity, the criminal offences will be the most central Event entity. Nevertheless, a criminal offence is rarely an impulse action. The event will most likely be planned in advance, followed by plans of escape routes and other actions. These are also key events in an investigation.

Events will also apply to the digital world. Pretty much all use of digital technology can be defined as events. Traffic between different services on the Internet are events, an email that is sent is an event, logging in to a service is an event and downloading of the child abuse material are events.

Exploitables

People who do anything fraudulent would like to try to hide their tracks. One of the reasons why criminals act on or through the internet is the possibility of anonymity (Bryant, 2014). Like other people they will still make mistakes and leave behind the traces that the police can follow. One source that will often provide information is e.g. when other people post a picture on a social medium of someone and maybe even tag that person by name, without the person's knowledge. Other people's openness will enable information about the target in an investigation to be accessed via a backdoor. (Ramwell, 2016).

Ramwell et.al (2016) points to two areas that are opening opportunities in an investigation. They refer to them as Laziness and Ego and call them Exploitables.

Laziness and Ego presents two reasons that information becomes available to the police when they have the knowledge and skills to find it. Laziness is related to the fact that people cannot bear to do everything that is required to keep information about themselves hidden on the internet, or they do not realise how much information is left open and what they need to do to prevent this (Ramwell et.al. 2016). Covering or hiding the tracks made on the internet would not only include those made by themselves, but also those others leave. As mentioned above, others will be able to post pictures and information about a person who is trying to stay

anonymous on the Internet and it requires a lot to make sure that no one else post information about you.

Ego is the other common pitfall in social media. People will consciously or subconsciously use social media to publish their emotions, thoughts and pictures of themselves. Furthermore, they will check in to various places or events they are on and they will like pages that fit with their interests and preferences. Most People, also criminals, wish to publish things about their life, especially things they have mastered in some way, possibly to brag or show off. This may also include the bragging of the criminal activities they are committing (Ramwell et.al. 2016). In 2017 Norwegian Broadcasting (NRK) made a report on travellers who ran human trafficking and prostitution in Bergen. They followed, among other things, the suspects open profiles on Facebook where they showed the large sums of money, expensive cars and photos from expensive parties, although they had no jobs or income. Information from different Facebook accounts was an essential part of the material journalist used to substantiate and document the case (NRK, 2017).

5.2.2.2 Validation

An important aspect of information retrieval from open sources are validation of information. As described over is Validated OSINT according to NATO: «...information (that) to which a very high degree of certainty can be attributed". The process of validation of information is further described below, but it is an important aspect to include in the mapping phase, namely how the information needed can be validated. This moment should be a factor both during the mapping phase and the planning phase. These phases will in effect overlap each other along the way and several key moments will emerge in more places. The actual conducting of validation will be carried out in the phases of collection, processing and analysis. Gibson (2016) points out that a significant moment in analysing the information is to move it from Open Source Intelligence (OSINT) to Validated Open Source Intelligence (V-OSINT).

The validation of open source information is not always easy to plan for. It must be done during collection, and partly through processing and analysis, but it is important to think about it when conducting the planning. It should be described as an important issue for the next phases. During planning one might include dividing information in what information is important to validate versus what information is not as important to validate. When collecting

information, it is important to mark what is validated and what is not validated when collected.

5.2.2.3 Value of information

"All information has value" is an argument one will quickly be faced with in a phase where one should map what information to search for. It is, of course, entirely true and hopeless to argue against. The truth is we wish for all the information that can point us in the direction of the target, e.g. uncovering a suspect. In an investigation it would also be desirable to gather as much relevant and detailed information as possible. The case is unfortunately often that the police do not have unlimited time and resources. One reason why digging journalists in many cases reveal much more information than the police might be that they make money by publishing their findings and can thus more easily utilise great resources. If a case is going to be read by a large number of people, the newspaper could use a lot of resources to retrieve all the information they need.

When limited time or resources result in some tracks not being followed up, it is important to consider what information is most likely to be found without too much use of resources and what information is most important to the case or project in process. Where the two criteria overlap, there is a good starting point for the job of collecting data. Where information is readily available but not necessarily as important to the case, it should be assessed whether the amount of information dictates if data should be collected. Where important information is believed to be difficult to find, it is necessary to consider other means of providing information. If there are no obvious alternative solutions and the information is important to bring the case forward or obtain evidence for the indictment, use of resources must be seen in conjunction with the nature of the matter. In a serious case like murder, abuse against children or other abusive crimes, the resources will be used while in the less serious cases use of resources is considered stricter.

5.2.3 Preparations

Before conducting the collection of Open Source Information, it is important to undertake thorough preparations. What preparations are needed will vary from case to case. No matter how the search and collection will be performed, it is important to think through which

preparations is necessary in this particular case. A list of possible topics to take into consideration is very useful for an investigator performing open source intelligence.

5.2.3.1 Equipment

A part of the preparations is considering what technical equipment and software you need. For some tasks you will need a new machine with no traces from the previous activity and a new internet connection, while others mission can be based on the reuse of equipment used if there is no need to search for information from places where the counterpart can monitor the activities. The consideration of equipment must be seen in the context of operation security.

Requirements for the information you are seeking and where you expect to find tit will describe what software you need. If there is a mission that requires retrieval of a lot of information or a broad search of many sources dictates that one should look for automated search solutions when manual searching can become too laborious. There are many sources of automated search queries that the individual official who works with OSINT should have a certain overview, knowledge and skills in. If there is a need to retrieve large amount of data from the Chatlogs, forum posts and the like, software is needed both to bring it out and structure it in retrospect.

In some missions, the use of specific Virtual machines like Tails[8] or Buscador[9] could be most useful. It is important to know the tools as well as their advantages and disadvantages so that this can be taken into consideration during preparations.

5.2.3.1.1 Hardware

Any device with Internet access can initially be used for Open Source Intelligence, but it does not necessarily provide high operational security. The device used must be able to run the software needed to be anonymous in those scenarios where anonymity is requested. Both mobile phones and computers will often have similar opportunities to run services such as TOR, VPN, etc. In order to blend in with the crowd it is necessary to use the same platform as the other users. If most users on a service use mobile phones, the use of a computer will stand

[8] https://tails.boum.org

[9] https://inteltechniques.com/buscador/index.html

out. The same relates to using a strange OS. If everyone uses a Windows computer, it will be noticeable if a user visits from a Linux machine.

5.2.3.1.2 Software

An early consideration when setting up a computer for the use of OSINT is access to a VPN[10]. The VPN client will encrypt and redirect all Internet traffic through the VPN-servers so the IP-address on the server side will be from the VPN-server (Bazzel, 2018). With a commercial VPN provider, it is usually possible to change country and server frequently, which can enable a new IP-address for each session.

What is beneficial on the one hand, can just a s easily be a drawback on the other hand. Some VPN services can be detected so that whoever monitors the server sees that traffic comes through a VPN. They cannot see who you are, but they can point it out as "suspicious" traffic.

Another software that may be useful to OSINT tasks is TOR (The Onion Router). It requires using a separate TOR browser. TOR uses a high degree of encryption and routes traffic through a Tor Circuit before redirecting to the internet[11]. The challenge is that it is easy to disclose that traffic comes from TOR, although it is almost impossible to see the origin. The IP address of all TOR nodes, including the exit nodes, are public and some servers are shutting down Traffic coming from TOR.

5.2.3.1.3 Accounts

Some collection of information may require that the investigator is logged in to various services, especially regarding social media, but also a number of other deep web services where the information is not indexed by search engines (Gibson, 2016). If a service depends on logging in, a consideration of what service is required, who registers and keep the login information and how information is displayed to other users must be done.

---

[10] A Virtual Private Network (VPN) extends a private network across a public network and enabled users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network. (Https://en.wikipedia.org/wiki/Virtual_private_network)

[11] https://www.torproject.org/about/overview.html.en

Social media like Facebook etc. will to a large degree connect different users to each other. By doing several searches on a person in Facebook, chances are imminent that both the searcher and the one being searched for will get tips about the other as "people you might know". If the investigator uses his private profile to seek information from the profile of a suspect, the suspect might see that the investigator emerges as «People you might know». Most criminals would be concerned by this.

If the information gathering will or might involve searching on social media platforms (and often it will) there is a need to make fake profiles. Most pages require an Email account to link the user account to. Then it is necessary to create an Email account that cannot be connected to the officer who operates the account (Bazzel, 2018). It should not pose a problem to use a fake account on social media several times to various searches. However, repeated searches against the same person from the same profile will also increase the likelihood that they will see the fake profile as «People you might know». It is important that the officer who performs the investigation is familiar with the challenges using a fake account in different platforms over time. In some cases, it has no practical significance that it is obvious that the account is disposed of by the police, but the important thing is that it does not link to a particular officer. It does not matter if the account is used over time, but a fraudulent account is usually a violation of the guidelines to the offering platform, e.g. Facebook[12]. As a result, the account may be closed, and it will probably not be possible to create a new account with the same e-mail address.

5.2.3.1.4 Operation Security

Operational security is an important part of the preparations. In many cases police officers and others has started searching for information and preserving information without considering operational security (Bazzel, 2018). Since the level of operational security will depend on the situation the preparations must take into consideration where the search and collection will be conducted. If we seek information from places where the counterpart monitors the server or otherwise can see traffic to the places you visit, then the level of operational security needs to be quite high.

[12] https://www.facebook.com/legal/terms

Any device connected to the Internet will leave large amounts of information. This information is among others, used by providers of search engines, online shops, social media, etc. to recognise you to show you advertising based on what the provider think suits your preferences. Any device that contacts a Web server will send a User Agent String to the servers it visits. This is for the Web server to properly display the web page based on the application you are using. A web page will appear differently on a mobile phone than on a web browser on a computer. This is based on information from the User Agent String.

An example of a User Agent String Is:
Mozilla/5.0 (Windows; U Windows NT 5.1; AGB
RV: 1.8.1.6) Gecko/20070725 Firefox2.0.0.6Eff-Primer p. 3)

In addition to the User Agent String, most commercial Web servers install cookies on your machine to recognise you the next time you come visit.

Web servers can also be set up with AWstats[13]. The server will then gather even more information about the user. The person who monitors the server can then, among other things, get information about which country the visitor is coming from, which OS they use, Screen Size, which search engine was used and which keywords were used to find the page.

There is a lot to consider when it comes to operational security and this must be taken into account in all aspects, like internet connection, anonymising services, machine setup, software, resources, etc. Operational security is basically about being anonymous. There are several ways of being anonymous, where one is to not be traceable. It may in some situations be suspicious in itself when, among a selection of users, one is the only counterpart unable to trace. The second is to blend into the crowd. If you appear as any user at that site, it is difficult to be "mocked" as a suspicious visitor. If you wish to blend into the crowd, you need to know how the other visitors "look" so that you can "dress up" to fit the crowd.

5.2.3.1.5 Internet connection

[13] https://awstats.sourceforge.io

There are obvious that a visit to a web server from a police computer using the agencies` internet access will give whoever controls the server a big red flag that beckons " Police". Yet this is totally unknown to many officers. When it comes to the Internet connection one can set up various options which provide higher anonymity.

An alternative is to get set up an external internet line from a private provider (not the provider the agency uses). You will then get an IP-address from this provider. When using a large provider in the country or area one is operating in, it is easier to blend in. Another issue to take into consideration is the exposure by using the same IP-address over time. Even the use of anonymisation services might cause an IP-Leakage and thus revealing ones genuine IP-address. It is fair to assume that as soon as an IP-address is suspected to belong to the police, the suspicion might spread in the criminal communities. With a router where it is possible to change the MAC-address, it will be possible to "trick" the ISP to assign new IP-address for a new connection.

Another solution is to use a 4G router. How it will work will probably be a little depending on the provider you choose and the country in which you operate. What is certain is that it depends on the fact that it is 4G coverage from the site of operations. The IP-addresses rotate more frequently in a 4G router, so the user will probably most times go online with a new IP-address.

5.2.3.2 Testing

A key moment during preparation is to test all the equipment needed in advance. It is imperative that this is done prior to the actual investigation as it should never be assumed that the equipment will respond as anticipated. This includes computers, software, internet connection and services. If the mission involves being fully or partially anonymous in the phase of retrieval, testing that the Internet connection and the anonymisation services used are actually providing anonymity is very important. If there are plans to use software to retrieve data from e.g. a chat log, the software must be tested on a chat log with the same structure. Everything has to be tested in the environment it will ultimately be used in so that the collection does not fail due to the environment being different than the test environment.

5.2.4 Chain of Custody

A part of the planning and preparations is also to protect the chain of custody It involves planning how to document an Audit Trail. All the way from the setup and preparation of the equipment and until the analysis is carried out, including all small tools, tasks, and to-dos should be documented accurately and precisely. Many tools and services will automatically log everything it does to secure an audit trail. This must also be tested in advance. If there are parts of the process you see is not being documented by the tool, you have to take that into account when using it.

All data that is collected must be preserved and stored in accordance with established methods and principles for securing digital evidence (Flaglien, 2018) and everything should be saved as read-only copies that are preserved untouched thereafter. Only copies of the preserved material should be used during analysis. The integrity of the preserved material can be safeguarded by using a cryptographic hashing[14], so the copies you work with can always be verified against the original. It is important in the documentation to be precise that it is the preserved material that has been verified and not the original on the Internet. In the case of data on the Internet, it will not always be possible to verify the original on the Web server prior to preserving it, so that changes to the collection process will also follow the downloaded and read-only copy.

It is especially important that Chain of Custody is Safeguarded in the collection phase, because in many cases there will not be possible to redo the process if there are any mistakes done. Therefore, it must be planned for how this should be done and how the integrity should be safeguarded.

[14] A cryptographic hash function is a non-reversible mathematical function that takes an arbitrary amount of data is input and returns a fixed-size string as output. The result is a hash value, and its mathematically infeasible to find two different files that create the same hash. (Flaglien, 2018)

5.3 Collection (Search, retrieval, and validation)

Today, everyone is in one way or another present on the Internet. Most people also use social media to various degrees as part of the interaction with other people. All actions using the Internet leaves digital footprints. These are these footprints that we wish to follow and collect data from. It is necessary to start with the information we already have. We might have a nick name, an IP-address, name of a forum, a school, an email address or similar. We have something to start off with to obtain the new information. The collection phase will to a large degree focus on finding pieces of information that can lead to new information, which in turn leads to new information. This can be called following the breadcrumbs.

Identifying what data is needed to meet the information requirement is first step in the process to determine the best source and method for obtaining this data. Just because data exists does not mean they are readily available. The required data can both be difficult to find and be presented in an unusual format. Common searches with search engines, is a task most police officers manage, but to perform power search will be outside of most ordinary investigators scope. In addition, a lot of information could reside on different forums etc. and anyone who will work with Open Source Intelligence must learn to search for information in more unfamiliar places like, among other places, online Chat-Services and file-sharing networks (Ferraro & Casey, 2005)

The process of finding information can both be done manually and by using software that make searches on different search criteria in many places automatically. Based on the defined requirements and planning an investigator will seek information where the information is most likely to be found. When using tools for automated search, much of the job will be associated with processing and validating data as the automated tools probably will preserve the information automatically. An important part of the collection is that it will constantly uncover new information to search for and things that have not been accounted for in the planning will also emerge. One must also be creative in finding new places to search and new ways of retrieving information.

A list of all potential available information about someone could be as comprehensive as a book. We have to narrow it down to what we need of information to achieve the result we aim for. To do that we need to have a systematic approach. There will always be some general

information that is needed regardless of the case, as well as there will be a lot of information that is relevant in one case, but not in another.

When gathering information from open sources on the internet, the amount of information may be overwhelming. What should have priority when collecting information may vary. The collection of data can be an extremely manual job of search on the web, search on social media platforms, forums, newsgroups, etc. to find pieces of information, such as a phone number, a user name, an identity, etc. which can be used further. The officer collecting information must always evaluate the use of manual search against the effects of automated tools. There will naturally be a big difference between officers who gather information from open sources on rare occasions in cases at work and those who have Open Source Intelligence as a main task in their work. In the latter group there will probably be considerably greater use of automated tools, because they do this so often that they locate these tools and learn how to use them, and in some cases, create them themselves.

When collecting information, it must be considered which tracks to follow, and which should be put aside or prioritised later. It is important to keep track of the leads that are not followed at the time, as they may become relevant at a later stage.

A good practice for preserving and documenting all information retrieved is highly important. The principle of Chain of Custody through an audit trail must be followed. In terms of digital evidence, they will be able to change quickly and if the audit trail is not followed, Chain of Custody can soon be broken. The more volatile the data, the more important it is to document thoroughly as it may prove difficult to verify later at a later stage

Preserving data which is to be used as evidence for prosecution must follow the principles, standards, and methods of computer forensics. «Computer Forensics is the scientific collection, examination, authentication, preservation and analysis of data held on or retrieved from computer storage media in such a way that the information can be used as evidence in a court of law» (Gottschalk, 2010, p. 141). During the collection it is not always possible to know if the information will be used as evidence, so all collection should be done according to these principles and standards.

It is also important to think of the order of volatility. Multiple sources are often searched at the same time or in the same process. It will differ how volatile the different data is, even on the internet. The principle of order of volatility describes that: «Prioritisation of the potential evidence source to be collected according to the volatility of the data» (Flaglien, 2018, p. 30). It is therefore important to prioritise preserving the most volatile data first, in as far as they are identified. In computer forensics it applies in the way that data stored on disk is less volatile than data stored in memory (Nist, 2006). On the Internet, data stored in news along with maps and satellite will be less volatile than data stored on forums, social media and newsgroups.

Preserving information from open sources on the Internet can be done in many ways, with different tools. There is no need to have access to expensive or advanced tools to secure information, but they can be useful when a lot of information is to be collected from different sources. How you choose to collect data depends on the information you are looking for. If it is appropriate to secure all information from the website of an organisation you are investigating it would be useful to preserve the entire site both with a form of "Print Screen" which shows how the Web pages look as well as the entire website with source code. In other cases, it might be sufficient to document only a single post on a forum with user name, content, and timestamp. In some situations, it can be preserved with a «Print Screen» and the URL in the address bar of the browser. What is important is preserving the data in such a way that the Principles of a forensically sound investigation is followed and that the proof will stand up in court.

During the collection, some information irrelevant to the specific case will collected. There might be information that may be relevant in other ongoing investigations, there might be evidence of criminal matters that are not reported or there might be information on possible criminal offences that are yet to be committed. This can be both personal information and other information and such information must be handled appropriately and according to the various countries' jurisdiction. Basically, the investigator must ensure that the personal information collected is linked to the case in such a way that it can be used further in the investigation (Ramwell et.al., 2016).

5.4 Processing

During Processing the collected data will be examined and prepared for further analysis. It is important to document every step during the process in order to safeguard the chain of custody. An examination of digital evidence will often involve restructuring, parsing and reprocessing of data to make it understandable for later analysis.

In case of collection of information from open sources on the Internet, this stage will involve reviewing the data collected to prepare for analysis. That might be to process the data to a common format so that data from different sources can be compared. It may also be to go through a preserved web page to collect data to be used further in the analysis phase. Basically, the processing is about converting the data into the required format for analysis, fuse it with other data sources, identify relevant data and begin the extraction and aggregation process. (Gibson, 2016)

Data will be retrieved in both structured and unstructured form. Structured data is data that is readily contained in databases with explanations for the different tables and the cell and the relationship between them. Unstructured data is the opposite where data does not emerge in structure with a model that describes content and relationships. Unstructured data can typically be Web pages, images, video, and other files that basically require a more manual review.

Natural Language Processing is a method of processing text from open sources. A large proportion of the data secured by Open Source Intelligence will be in a text format and are thus available for indexing and search. What may be a challenge, especially with the securing of data from forums, social media and equivalent, is that users could write using regional colloquialisms or dialects not recognised in written form; it can be riddled with typos and in addition, usernames and nicknames will have to be considered. Here, too, the investigator will benefit of learning the use of automated tools to make the job easier.

Identifying entities is an important part of processing. Entities are important in the analysis process and therefore crucial to extract. The analysis can reveal that several entities can be fused as username and nickname can be connected to a particular person or it may uncover that other seemingly lone standing entities in the material appear to be the same.

Modelling is another method of processing. Often, text-based data will be put in a context. There may be a Chat or a forum thread where putting messages in context is important to give any meaning. On social media, often a post will be in response to another post. Then the different records must be put in context so that links can be established.

5.4.1 Validation

Validation is an ongoing process from collecting to processing to analysis. Validating must be done as an integral part of the retrieval, because it is while searching and securing information it is most natural to assess the source of the information, check for information that can verify the source's credibility and search to verify the information from other sources. However, validating all information during collection is not possible, so the validation process continues during processing. Gibson (2016) places validation of the information purely as part of the processing.

Validation of the information found via the Open Source Intelligence can be a challenge. It will largely depend on the source of the information and the assessment of the source is therefore a central part of the validation of information from open sources.

As mentioned above, an important part of Open Source Intelligence is to move information from OSINT to V-OSINT. By establishing and adopting methods to prioritise, assess credibility and confirm sources, information and intelligence we increase the chance that what is presented has high reliability, precision and value. This creates a better foundation for making decisions based on Intelligence (Gibson et.al., 2016).

Assessing credibility may be a difficult task, therefore assessing the source is paramount. NATO uses the following recommendations to assess credibility of a source:

1. *The authority of the source,*
2. *The accuracy of the source,*
3. *The objectivity of the source,*
4. *The currency of the source and*
5. *The Coverage of The Source*

(Gibson et.al. 2016 p. 106)

Information from renowned media houses will have a significant higher credibility than the simple records on social media, like Facebook, Twitter, Instagram etc. Accuracy and objectivity will often be higher in those who are considered to be credible sources. If a source does not provide a writer and date for publishing information, it gives little credibility. Then the information is hard to verify. Information from social media, forum etc. can be much more difficult to verify as the content has been generated by users without some form of quality or editorial control. An individual may publish in these forums, but the published item may only be significant due to its existence. However, when compared and cross-referenced with other information, the published item can potentially be verified or disproved. Information about people's locations on from social media may be confirmed from other sources, e.g. travel companies or other countries' authorities by passport control. Fusion of OSINT And Non-OSINT Data can also be used for validation of information, but such validation is usually done in the analysis phase.

Different intelligence agencies often use different methods to classify credibility on scales from four, five, six and more values. NATO has its method, the U.S. Army has its method And UK Police has its method (Gibson et.al. 2016). Although the various agencies use different methods of assessing credibility of information, the interesting question in terms of methodology is; how varied is the criteria used to place information on a scale from four to six and so on? It should not be based on a subjective consideration. It is important to, as far as possible, use objective criteria to assess the credibility of the information. If the model does not provide clear guidance on how validation should be carried out and which objective criteria should underpin the assessment, such models will be of low value.

Verification from other, independent sources will be an objective assessment of information. All sources will also need to meet same assessment criteria as the original information. The problem may otherwise be that the same error in information is spread by many because of its nature. An event that is shared by many on social media will not automatically get high credibility, even though it is spread by many. Information about actor Eddie Murphy's death reached large part of the world even though it was false[15].

[15] http://www.nbcnews.com/id/46257734/ns/technology_and_science-security/t/eddie-murphy-alive-despite-twitter-death-hoax/#.W-LdLKeDpBw

News articles can more easily be validated because there will often be other news media reporting the same. An important checkpoint is to make sure that not everyone refers to the same source. Then the possibility that it is incorrect would be as high as if only a single media house reported it, but most serious media houses will seek to verify the information from multiple sources or take reservations if that has not been possible.

It is not always possible to verify the information collected from open sources. In the cases where verifying information is not possible, it is important that it is documented in such a way that lack of verification is clear and easy to see. Information that cannot be verified will not be acceptable as solid evidence, but it is circumstantial and can therefore support a hypothesis. In a proactive investigation, the requirements for verification will often be lower to the extent that information will be used as basis for police response, not used as evidence for prosecution.

5.5 Analysis

Information will not become intelligence without analysis. The ability to analyse the information is what distinguishes basic OSINT from excellent OSINT (Hribar et al. 2014, as referred to in Gibson et.al. 2016 S. 95). There are many forms of analysis that can be carried out in Open Source Intelligence and there are various tools that can be of help.

Analysis is about evaluating and compiling information to support or reject the hypotheses. An essential part of an analysis is to see information in context. It is important to structure the information to enable a clearer view of information that would otherwise be hard to see.
A standardisation of information in fixed categories makes the analysis easier. With the POLE data model (TechUK, 2014) as a starting point, the analysis identifies the various entities in the retrieved material. A key value for the investigation is to identify relationships between different people, objects, locations and events in a case.

The entities in a case will be associated to one or more other entities. When we collect information, the relationship between these entities is often among the most important information. It does not mean that the information about the connections is known, but there must be a tie. Relationships will give clues about links between the various pieces of information. This can be analysed by inserting it into a matrix for later export to a visualisation tool such as i2 Analyst's Notebook[16] or Maltego[17]. Such a matrix can be filled in continuously through the analysis and will also display the fields where information is missing.

Entity	Entity Type	Description	Relation	Verified	Entity	Entity Type	Description
John Doe	Person	121265-3452	Owner	Yes	Car	Object	ABC1234DE
Robbery	Event	Case: 12345678	Observed	No	Car	Object	ABC1234DE
John Doe	Person	121265-3452	Suspect	No	Robbery	Event	Case: 12345678

Fig. 2. Relationship Matrix

[16] https://www.ibm.com/us-en/marketplace/analysts-notebook

[17] https://www.paterva.com/web7/

When exporting to a visualisation tool such as Analyst's Notebook one can choose to get verified links as a whole line while unverified links are dotted lines.

The collected data can be further analysed, among other things, by using text analysis, network analysis, location analysis, and time analysis. Usually, several forms of analysis are performed simultaneously but presented differently. A network map can visualise who has a relationship to whom and to what extent the relationship is verified as well as the strength (e.g. the amount of contact) in the relationship. A network map or network analysis is often used for to find links between different person entities, see what kind of links they have and review and validate the links. A network map can also be used to visualise relationship between people, organisations, vehicles, and other entities. In a network map, one can also make a selection of central links to see correlations of various types of connections between entities. Both Analyst's Notebook and Maltego can automate the job of network analysis.

A timeline can visualise information in a way that makes it easier to see the sequence in which events have occurred and at what interval. It can provide information that was not discovered when the information was brought in and it can also display areas where the data is deficient, e.g. that an event is missing to confirm the hypothesis. Finding information about this event will then become the new task.

Data can also be presented in maps to indicate where different events have occurred or location of different entities. It can be where different people live or stay, there may be locations for different events and there may be a juxtaposition of different information from various sources.

When gathering information, a compilation and analysis of information from multiple sources is part of the task. In the intelligence doctrine of the Norwegian Police, this is described as multi-source analysis (Pod, 2014). This part of the analysis is about assembling information from various sources and looking at correlations that are not revealed by analysing each individual source separately. The various entities in the collected material will often appear in several types of data and the data may have a different validity. The challenge may be to merge the various identities of an entity into a unique identity. There could be a Facebook profile that is verified to belong to a particular person (e.g. a suspect) followed by a couple of identities from different forums and they may all belong to the same person. How can these be

merged when there is no certainty that they are the same entity? Here, a new hypothesis develops that needs to be verified or be falsified. If verification or falsification of the hypothesis is unsuccessful, the uncertainty must arise in the presentation of the analysis. On the other hand, this merger of identities will be a good tool if it can be verified that the identities belong to the same entity. Suddenly there is much more information about the entity.

Day, Gibson & Ramwell (2016) presents a similar process of fusion of OSINT and non-OSINT Data. The process describes merging information coming from OSINT with information from closed sources. The closed sources may be information from e.g. Police records, Telecommunication Records, Population registers, Financial records, Medical Records and other information that is not publicly available. Traditionally, the police have largely utilised these closed sources and informants, to provide intelligence. The use of open source information has not been seen as a potential until recently. When police officers see the potential of both OSINT and non-OSINT data, they will have a broader range of sources to retrieve information from. There is also more knowledge to gain by consolidating several different identities into one with the information contained in both OSINT data and non-OSINT data.

Analysis is often a repetitive process that generates new information, which in turn generates new tasks. The new tasks will generate the collection of new data for processing which then will be analysed. It can therefore be seen as a circular process from collection to processing to analysis and on to new collection, new processing and new analysis. This circular process is in progress until the required information is gathered and analysed and the result can be disseminated.

5.6 Distribution and evaluation

5.6.1 Documentation and keeping an audit trail

An online investigation or gathering of information from open sources will involve the collection of volatile data. Not all the information found may be recreated and it will in many cases be difficult to reproduce the process and the findings, simply because the data are no longer there. Therefore, it is important when working with volatile data, whether it is Live data Forensics, Online Investigations or Open Source Intelligence, that continuous documentation on how information is found, retrieved, processed and analysed is conducted. This can be done manually or automatically. Some tools such as Fireshot will ensure time stamp when taking screenshot of a Web page.

The importance of documentation and an audit trial repeats in many standards and guidelines for the investigation of digital evidence. In RFC 3227, the importance of "documenting every step" is pointed out when preserving digital evidence and also how important it is to safeguard «the Chain of Custody» (The Internet Society, 2002 p. 6-7). ACPO Guidelines also stress the importance of an audit trail and documentation where principle 3 is described as: «An audit trail or other record of all processes applied to digital evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result» (ACPO, 2012, p. 6). Also, NIST SP 800-86 clarifies the importance of Documentation: «*The documentation allows other analysts to repeat the process later if needed*» (NIST, 2006, p. 28).

Documentation is a consistent process. Even before starting to search for information, the time, location, hardware and software used should be documented and this should be consistent throughout the entire process. Finally, this should result in a report.

5.6.2 Report

It is important to take into consideration that the report must be transparent and understandable to the reader. The first thing to think about is who the audience is (the recipient) of the report. If the report is to be read by a prosecution lawyer or be used as evidence in court, technical terms should be avoided as far as possible. The essential technical terms must be described and a summary of the essence of the process should be provided.

Flaglien presents the typical points for a report when examining digital evidence:

- Roles and tasks for the investigation,
- Executive summary of all information sources and evidence,
- The forensic acquisition and analysis, which reflect chain of custody and evidence integrity,
- Visualisation and diagrams,
- Images and screen shots,
- Information that supports repeatability and reproducibility of the analysis.
- Tools used, and
- Findings.

(Flaglien, 2018, p. 46)

Bazzel recommends that the findings and the conclusion are presented in an executive summary and that information about the sources, process and evidence will be presented later in the report (Bazzel, 2018).

Bazzel describes that his reports tend to contain the following points:

- Executive summary: One-page synopsis of vital evidence
- Suspect Details; Specific data such as all personal identifiers, user names etc
- Narrative report: Detailed findings with references to digital evidence and summaries
- Summary report: One-page summary of facts and need for further work.
- Digital Evidence: A DVD or thumb drive that contains all screen captures and files.

(Bazzel, 2018, p. 456).

During presentation of the analysis phase various forms of analysis were described that can visualise the information in a very transparent way, like network maps, timelines etc. It would be advantageous to incorporate these in the report. It will be much easier for the reader of the report to understand the result by seeing the visual representations rather than relying purely on text. As Flaglien write, "*A wall of text is no good for anyone*" (Flaglien, 2018, p. 46). A report that is transparently written with good visual effects that display key evidence will be

much more useful for whoever gets the report and will use the information further, e.g. as basis for the indictment.

The more detailed process that is performed manually or with the use of tools can be presented as an attachment or under a separate section of the report. This part can reveal what searches and findings has been done. The entire detailed process of searches, operators, URL's visited etc. can be included there. This information will be mostly of interest to the one who will validate the process. Many of the tools which can be used in a digital investigation generate their own reports. These will also be natural to add as attachments to the report whilst including important issues in the summary

The documented Chain of Custody is the glue that keeps the methodology and processes together. The report will safeguard the principle of Chain of Custody so that no faulty documentation of tasks or discoveries make it possible to ask questions of the validity and integrity of the evidence in court.

There are many examples and templates for how such reports can be structured, and there is no definitive answer of which best way is. The main points must be that the report has a transparent and easily-understandable summary that addresses the most important findings and that all tasks are well documented.

An important part of such a report will often be recommendations for further action or other investigating tasks. Here, it is natural to distinguish on reports that are to be part of the evidence for an indictment in a criminal case and reports to be used by the management of the agency as decision-making.

In a criminal case report, recommendations and measures will be reserved for temporary or internal reports. The final report that is used as part of the basis of indictment will be a final report where all recommendations are conducted. A summary of the process and results will be the central part of a criminal case report.

In an intelligence process to retrieve the necessary information to decide on response to prevent crime, the recommendations and measures will be the central section of the report.

5.6.3 Distribution

The result, described in the report or presented otherwise, must be distributed to those in need of the information. In a criminal case there will be investigation management and prosecuting attorney. Furthermore, completed reports in the criminal case shall be distributed to defence attorneys and other parties.

An intelligence product must be made available to those who need the information. It could be the officers patrolling in the area where the problem exists, it could be anyone in a project working towards a type of crime, there would often be management, there could also be external collaborators like child welfare, social office etc. It is a managerial responsibility to represent the correct and appropriate distribution of the knowledge.

5.6.4 Evaluation

Evaluation of the process is important to assess whether the goal is achieved. The result should be weighed against the hypotheses and the mission. Is the required information retrieved? If not, why not? Have the proper searches been conducted? Were the right tools used and has anonymity been safeguarded? This are key questions to ask in an evaluation. It is easy to forget the evaluation, especially when the goal has been reached and the mission is resolved. Nevertheless, it is important to see the processes in the light of what the requirements was.

Evaluation can be done in many ways. It is most important to evaluate the result up against the initial requirements. Have the initial questions been answered? Moreover, it would be important to consider if the result generate new questions that need to be answered. It is also important learn from the process for later use, both what was successful and what went wrong. At last it would be useful to evaluate use of resources. Is time spent on the right issues? Maybe a disproportionate amount of resources was spent on something that did not lead forward? Should we have finished a few searches earlier?

6 A Model for Open Source Intelligence Methodology

The presented methodology contains of six phases. The base of the investigation is that a phenomenon or organisation is developing or has developed, or a criminal case is reported. Based on this, some hypotheses are developed that show some requirements. From the requirements, a strategy is developed which further forms the basis for the planning of the process. Upon completion of the process information is collected and processed for further analysis. The final product is a presentation of the results made in a report or equivalent.

The same way the intelligence circle goes in a circle from “Direction” to “Dissemination”, the OSINT Process goes in a circle from "Requirements” to “Dissemination and evaluation”, where the product is not only to be distributed, but is evaluated against the original hypotheses. Here, one will, based on the result, set new hypotheses that give a new direction and the process starts all over again. As the model in figure 3 Shows the process could go in a circle.

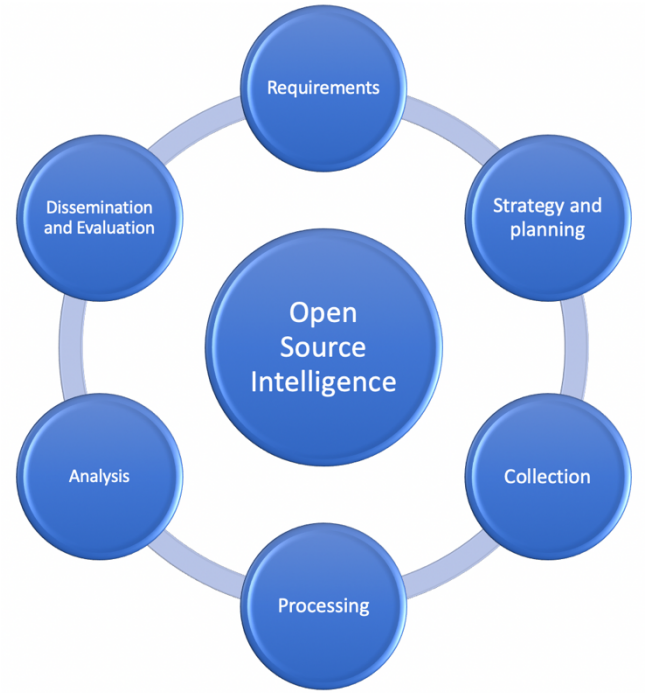

Fig. 3. The OSINT Circle

The goal of Open Source Intelligence is to assist in verifying or falsifying those hypotheses established at the outset. Information may also appear along which forces new hypotheses to be formulated. In the last phase, recommendations and measures in the report will describe if there is a need to start a new process

The practical work of police officers who perform Open Source Intelligence, whether as a specialist or as a part of everyday work as an investigator, will not be as rigid as the model presented. Still, a clarification of which stage in the process the tasks are done governs who makes decisions and who acts. When management has given the assignment based on hypothesis operationalised in requirements and strategy, the operator will not set aside the hypotheses, create new or define new requirements. The operator's role is to deliver a product based on the assignment. Management will potentially set new hypotheses if the operator’s presentation and results provide a reason to do so.

The practical implementation of Open Source Intelligence will partially be a circular process as the result of some processes will lead to new tasks in other processes. Intelligence from the analysis will lead to a demand for new information, which needs to be processed and analysed. As a result, there will also be a circular process inside the OSINT model from which it goes from collecting to processing to analysis and back to new collection.

The model for Open Source Intelligence Methodology may be presented in two ways. The first is circled from Requirements to Dissemination. For management, this model will be most useful as it shows the phases in sequence. For whoever performs Open Source Intelligence however, the process will start with the Requirements, Strategy and Planning that defines the direction. After planning and preparing, the actual process will circle between the collection, processing and analysis until the final result will be presented and distributed.

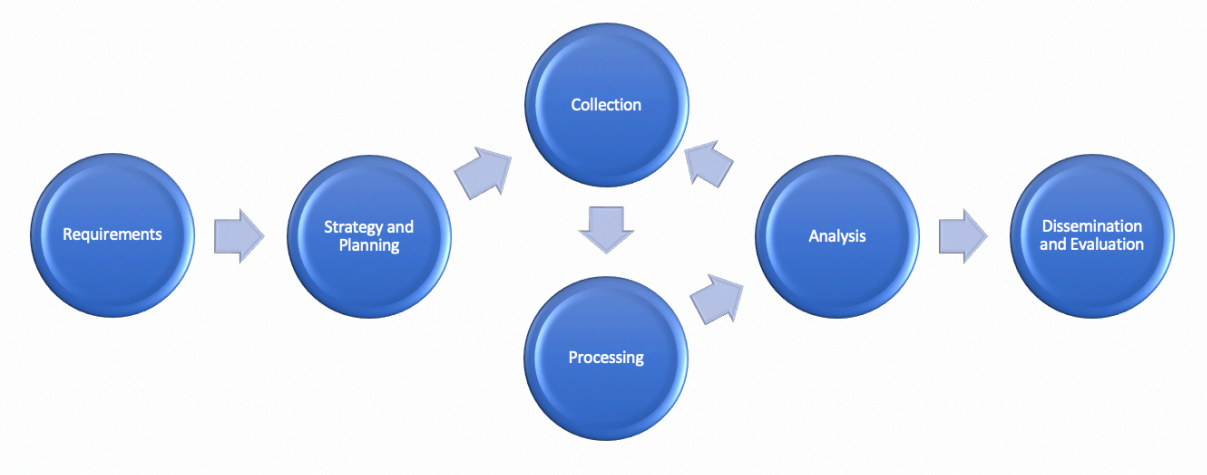

Fig. 4. The OSINT Process from execution level

In the Intelligence doctrine of the Norwegian Police it is defined whom has what responsibility in the intelligence process (POD 2014). The same responsibilities will be transferable to the OSINT Process. Figure 5 shows the different phases of the OSINT process with the corresponding level of conducting and responsibility from the intelligence process.

Phase in the Intelligence Process	**Phase in the OSINT process**	**Executing**	**Responsible**
Prepare, lead and Prioritize	Requirements	Executive Officer	Executive Officer
Planning	Strategy and planning	Chief of Intelligence Operator	Chief of Intelligence
Collection	Collection	Operator	Chief of Intelligence
Processing	Processing	Operator	Head of Operations
Analysis	Analysis	Operator or Analysist	Head of Analysis
Dissemination	Dissemination and Evaluation	Operator and Chief of Intelligence	Chief of Intelligence

Fig. 5. Responsibility and execution compared to the intelligence process and OSINT Process

When looking at what level has responsibility and what level executes the different phases of the process it is natural that it will become a circle among Operator-level tasks, before the result is delivered to the management. Figure 6 shows the same model as figure 4, but with description of the level of execution.

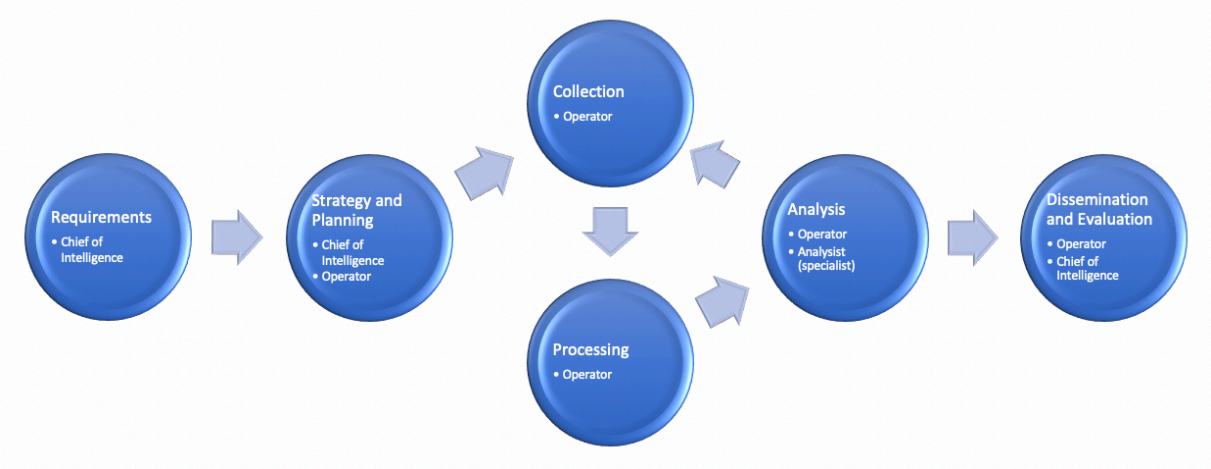

Fig. 6. OSINT the process from the level of execution with description.

6.1 Comparing models

There are various models that describe the process of investigating digital evidence. They may look different and may be adapted to various forms of digital investigation, but in essence they contain many of the same elements. The process described in this dissertation is compiled here with a few other models to see the common denominators and differences.

Flaglien (2018, p. 15) presents the Forensic Process as a process for the investigation of digital evidence. The Forensic Process defines a structured investigation of digital evidence from any source liable to store data. The process Flaglien presents should be universal in terms that it can be used on investigating any digital evidence regardless of the case in question. The process consists of five phases:

- Identification

The first step in the process is to identify the digital evidence and where they are stored.

- Collection
  Preservation of the information must adhere to the principles and standards applicable to preserving digital evidence to safeguard authenticity, integrity and reliability.
- Examination
  During the examinerships process, the collected data should be examined and prepared for further analysis. An examination of digital evidence will often involve Restructuring, Parsing and Reprocessing of Raw Data to make it understandable for later analysis.
- Analysis
  The collected data must be analysed in order to make the content meaningful and to make sure that the digital information can be used as evidence
- Presentation
  The result is rendered in a report where Chain of Custody is documented, and all processes, tools and results are described.

Hassan and Hijazi (Hassan & Hijazi 2018 S. 343) presents their model of a five-step OSINT process:

- Identifying the sources
  Identify the sources to collect the data
- Harvest the data
  Use of tools and techniques to gather data
- Process and verify data
  Process the gathered data and verify uncertain data from other sources
- Analyse the data
  Analyse the data to find connections to complete the picture
- Deliver the results
  Present the result

If the various models are inserted into a matrix, significant similarities can be seen here in figure 7:

OSINT Process	**Forensic Process City Flaglien**	**OSINT Process by Hassan**
Requirements		
Strategy and planning		
Collection	Identification Collection	Identify the sources Harvest the data
Processing	Examination	Process and verify the data
Analysis	Analysis	Analyse the Data
Dissemination and evaluation	Presentation	Deliver the Result

Fig. 7. Comparison of models for the investigation of digital evidence

It can be seen that input values describing what is being searched for are missing in both models. It can be argued that there is a case or an assignment at the core, but this is missing from the model. Their models also lack strategy and planning. Having a plan before getting started with searching and ensuring digital evidence is very important and cannot be overlooked. Their models are to a greater extent designed for the executing party which will perform the practical part of the process. I would suggest that as a model of a process for investigating digital evidence it falls short due to the lack of requirements, strategy and planning.

7 Challenges

Open Source Intelligence is a large but not highly utilised source of intelligence, at least for the general investigator. With the benefits provided by the Open Source Intelligence it also comes hand in hand with some challenges. Hassan And Hijazi (Hassan & Hijazi 2018) cites three special challenges for OSINT, which can to some extent also be linked to other digital investigations.

1. Sheer Volumes of Data:
   Collection of information from open sources may provide huge amounts of data to be analysed. Fortunately, there are tools to automate parts of the analysis job, but it is still a large job, especially with unstructured data, to structure those in a format that fits with the analytics tool.
2. Reliability of Sources:
   It can be a comprehensive job to verify all the information collected. Not all information is verified and must be used with care. In many cases, when the amount of data becomes extensive, it will be necessary to sort out what information to verify, because it would be too much work to verify everything.
3. Human Efforts:
   Although the analysis job can be partially automated, it will require people to review the results as well as validation of the information. Additionally, the information should be merged with information from other sources. This will take time and demand human resources.

(Hassan & Hijazi 2018 S. 16-17)

Another challenge with Open Source Intelligence is languages. For those who only work nationally, they may find that most of the information is presented in their own language or English, but even nationally, the investigators will investigate cases with foreign criminals that communicate in other languages. As mentioned in the example from NRK, all the text on Facebook was in Romanian. Those who work with cross-border or international crime will find that language quickly becomes a barrier. Around 50 – 80% of Open Source Information is in languages other than English (Ramwell et.al., 2016).

8 Summary

Open Source Intelligence will become an increasingly larger and more integrated part of the everyday life of investigators worldwide, but also military intelligence, private companies and journalists will increasingly use the potential that lies there. For the police, the key issue will be to build competence, establishing good processes and knowledge of the potential of OSINT for the entire hierarchy from the supreme leader to the patrolling officer. Furthermore, for the police and others who operate Open Source Intelligence it will be important to see opportunities to achieve a greater degree of automation of acquiring, processing and analysing data, as the biggest challenge may rapidly become to analyse all the information that can be collected. This will require new ways of thinking and new tools to work with.

It is already being argued today that Open Source Intelligence is the most powerful tool an investigator has, above all other sources and all other intelligence (Gibson, 2016). The tool has the potential to be even more powerful with our ever-increasing use of internet-based services.

Although the technical aspects of Open Source Intelligence are the most important in order to manage to effectively find, collect, process and analyse data from open sources on the internet it is essential to have a solid methodology as foundation that controls the process through the various phases. As shown above, there are several models for this process and similar processes, but several of them tend to lack significant phases or are adapted to one or the other level of responsibility. This methodology should be a guideline both for the management who require the information and enables the process and the officer that shall conduct the actual collection, processing and analysis.

This dissertation has intended to collect existing knowledge, evaluate it and use it to establish a methodology that is as comprehensive as it is simple and as it is technology neutral, it will hopefully be equally as relevant in a few years' time as it is today.

References

Akhgar, B. (2016). OSINT as an Integral Part of the National Security Apparatus. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 3-9). Cham: Springer

Association of Chief police officers. (ACPO). (2012). *ACPO Good Practice Guide for Digital evidence.*

Bazzel, M. (2018). *Open Source Intelligence Techniques: Resources for searching and analyzing online information.* CreateSpace Independent Publishing Platform

Bjerke, O. T. & Fasing, I. A. (2018). *Investigation: Principles, methods and practices.* [Investigation: Principles, methods and practice] Bergen: Fiction the publisher

Bryant, R. (2014). Digital Crime. In Bryant, R., & Bryant, S. (Ed.). *Policing Digital Crime.* (Pp. 1-42). Farnham: Ashgate.

Bryant, R. & Kennedy, I. (2014). Investigating Digital crime. In Bryant, R., & Bryant, S. (Ed.). *Police digital crime.* (Pp. 123-145). Farnham: Ashgate.

Center for Security Studies CSS. (2008). *Open Source Intelligence: A Strategic Enabler of National Security.* Retrieved from Http://www.css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSS-Analyses-32.pdf

Day, T., Gibson, H. & Ramwell, S. (2016). Fusion of OSINT and non-OSINT Data. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 133-152). Cham: Springer

Dilijonaite, A. (2018). Digital Forensic readiness. In Årnes, A (Ed.). *Digital Forensics*. Hoboken, (Pp. 117-146). Nj: Wiley

European Convention on Human Rights. (EHCR).

FBI. (2004, Nov. 15). Speech by Director Robert S. Mueller. Retrieved from Https://archives.fbi.gov/archives/news/speeches/the-fbi-improving-intelligence-for-a-safer-america.

Ferraro, M., Casey, E., & McGrath, M. (2005). *Investigating Child Exploitation and pornography: the Internet, the law and forensic science.* Amsterdam: Elsevier.

Flaglien, A. O. (2018). The Digital Forensic Process. In Årnes, A (Ed.). *Digital Forensics*. (pp. 13-49). Hoboken, NJ: Wiley

Gibson, H. (2016). Acquisition and preparation of Data for OSINT. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 69-93). Cham: Springer

Gibson, H., Ramwell, S. & Day, T. (2016). Analysis, Interpretation and validation of Open Source Data. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 95-110). Cham: Springer

Gottschalk, P. (2010). *Investigation and Prevention of Financial crime: Knowledge management, intelligence strategy and executive leadership.* Farnham: Gower

Hassan, N. & Hijazi, R. (2018). *Open source intelligence methods and tools: A Practical Guide to online Intelligence.* New York, NY: Apress

Kleiven M. E. (2005). *Where's the intelligence in the National intelligence model?* (Masters dissertation). Institute of Criminal Justice Studies. University of Portsmouth.

National Institute of Standards and Technology (NIST). (2006). *Guide to Integrating Forensic techniques into Incident Response: SP 800-86.* [Special Publication]

National Police Directorate POD. (2014). *Etterretningsdoktrine for the police.* [Intelligence Doctrine for The Police]. Norway

NATO. (2001). *Open Source Intelligence Handbook.*

NATO. (2002). *Open Source Intelligence Reader.*

Nhàn, J. & Huey, L. (2012). 'We don't have the laser beams and stuff like that': Police investigations as low-tech work in a high-tech world. In Leman-Langlois, S. (Ed.). *Technocrime, police, and surveillance (Routledge frontiers of Criminal Justice).* (pp. 79-90). New York, NY: Routledge.

Norwegian Broadcasting. (2017). *Sex, Dope and Beggar Cup.* (sex, drugs, and Beggars Cup). RET Rei at from https://www.nrk.no/dokumentar/xl/sex_-dop-og-tiggerkopp-1.13463802

Ramwell, S., Day, T. & Gibson, H. (2016). Use cases and Best practice for LEAs. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 189-211). Cham: Springer

Rogers, C. (2012). Intelligence Gathering and police Systems. In Awan, I., & Blakemore, B. (Ed.). *Police the cyber hating, cyber threats and cyber terrorism.* (pp. 129-148). Farnham: Ashgate.

Sampson, F. (2016). Following the breadcrumbs: Using Open Source Intelligence as evidence in Criminal proceedings. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 295-304). Cham: Springer

Schaurer, F. & Störger, J. (2013). *The Evolution of Open Source Intelligence (OSINT).* RET Rei at From Https://www.afio.com/publications/Schauer_Storger_Evo_of_OSINT_WINTERSPRING2013.pdf

Staniforth, A. (2016). Police Use of Open Source Intelligence: The longer Arm of Law. In Akhgar, B., Bayerl, P. S. & Sampson, F. (Ed.). *Open Source Intelligence Investigation: From strategy to implementation.* (pp. 21-31). Cham: Springer

Sunde, I. M. (2018). Cybercrime Law. In Årnes, A (Ed.). *Digital Forensics*. (pp. 51-75). Hoboken, NJ: Wiley

TechUK. (2014). *Breaking down barriers.* [Report]. Retrieved from Https://www.techuk.org/insights/reports/item/2302-techuk-launches-breaking-down-barriers-report

The Intelligencer. (2017). *Stop the violence! Teens ' obsession with sharing fight videos on social media an alarming trend.* Retrieved from: Http://www.theintell.com/6067fa3e-8b35-5ba8-94cf-c6106b1d6a1b.html

The Internet Society. (2002). Request for Comment (RFC) 3227.

Årnes, A (Ed.) (2018). Introduction. In Årnes, A (Ed.). *Digital Forensics*. (Pp. 1-11). Hoboken, Nj: Wiley