The power of OSINT

By Arno Reuser

Reuser's
Information
Services

Introduction

In 2017, a cyclist in Leiden got an argument with a car driver. In the heated debate, the cyclist got violent and threw the pregnant cardriver into the river De Vliet. Clever use of osint quickly learned the name and address of the suspect. How did they find the suspect? How can the clever use of OSINT help getting the right intelligence to find answers?

Summary

OSINT is an integrated, collaborative methodology to retrieve a balanced, representative and validated set of the best possible information from open sources to - after careful scrutinization and analysis - produce actionable intelligence. Intelligence that can be used to support decision makers in managing change. OSINT is used in preparation of peace-keeping missions, in international conflict and security studies, in market intelligence to explore new market opportunities, by law enforcement to find suspects, by banks to vet new customers, and more. The private sector uses OSINT for almost anything that requires good information. Journalists, historians, researches alike. Below we describe what OSINT is, how it works and a few practical examples of applying OSINT.

OSINT is widely misunderstood as a bunch of technical tricks in support of cyber operations or something. It is not. In-depth knowledge of open sources, in-depth knowledge of search systems and search languages are the minimum required for OSINT operations. Running the RIS OSINT Intelligence Cycle will produce powerful intelligence reports that are true force multipliers. The OSINT system described was developed since 1990 when starting to establish the OSINT branch for the Dutch Defence Intelligence & Security Service. The system became since then known as Arno's OSINT Methodology.

How does OSINT make a difference?

Clever use of Open Source Intelligence tools and techniques can be a true force multiplier for any intelligence operation. Knowing how to find relevant and timely information in open sources and how to analyse that to get actionable intelligence can lead to quick decisions at low cost. But only when done by professionals.

In the above example, the cyclist and the cardriver were involved in a minor car accident at a narrow road next to the beautiful river De Vliet. A road often used by cyclists for relaxation and sport alike. Both got involved in a heated argument. At a certain point the cyclist got violent and threw the pregnant car driver into the river De Vliet, leaving her in despair. Clearly a thing for the police to solve.

How did they do that using OSINT?

Police were quick to act. The OSINT department first did some target profiling. What do cyclists do? How do they behave? What is a general characteristic of a cyclist? That one is not too difficult. They like to keep a record of their progress and their achievements. How do they do that? By using a bicycle computer that keeps track of essential data like distance, speed, averages, energy consumption, heart rate, blood pressure, and the like. Some of the more advanced also keep track of progress in training, and the exact route the cyclist has driven, dates, times, if you have an online account that is.

One plus one sometimes leads to more than two. And that is certainly the case in OSINT. The next question was, which commercially available bicycle computers keep track of training progress data online. It turns out that there were only four of those.

How many of these keep the progress data of their customer cyclist data online? All four did. So, it may be a bit of work but nevertheless, the next question was, of those four online bicycle computer databases, how many participants were cycling at the date and time of the incident near river De Vliet? Guess what, only one. Name included. Bingo! Now go to social media to check the name and find an address. Got you!

What is OSINT?

Everybody needs information. Everybody needs good information.

Information that is timely, reliable, actionable and validated. Information that can be used to take decisions. Decisions to yes or no deploy troops on a mission, or decisions to yes or no enter a new business market, or decisions to start a new research project. Law enforcement, financial institutes, international missions, strategic analysts, journalists, (inter) national governments, intelligence services (competitive or general or defence), students, researchers, international missions, and many more, all need a proper foundation or a good information profile to take decisions.

And all of them discover the power of OSINT.

Once properly applied, OSINT can be a true force multiplier once certain conditions are met: A very good understanding of the organisation of the global information landscape, a very good knowledge of open sources (where to find what), very good skills in searching databases and query logic, and obviously a very good understanding of the initial requirements.

In short, OSINT is creating a as perfect as possible match between the supply side and the demand side of information.

OSINT is a methodology developed by when acting head of the Open Source Intelligence Branch of the diss. Back in 1990 when being invited by the diss establish their OSINT branch, it became clear that there was no such thing as OSINT in Europe yet. Everything had to be developed from the ground up. Gradually, an OSINT methodology was developed, a system of getting timely, validated and reliable information in the correct time and correct format to the customer. In this case, strategic analysts.

The official definition is:

The object of OSINT, the information being collected and analysed, is called osinf. The official definition is:

There are some restrictions. Legal basically means no hacking, no ’computer network exploitation’, no password cracking etc. All must be done in a legal way. Ethical means that some open sources happen to be in the open domain that are clearly not intended to be there. You may realise that considering the amount of open sources available, the in-depth professional OSINT research techniques being used, the OSINTian may find information that was not supposed to be found from OSINF. That information is not considered OSINF and thus not part of OSINT. Examples are USB sticks found in the back of a taxi cab, information released by accident, or Wikileaks. That information was stolen, editted and manipulated before being published on the Net. It is therefore not part of OSINT. No misunderstandings here, the information is interesting and will be used (probably) in an intelligence context, but it is not pure OSINF.

How much information is there out there anyway?

A lot.

Very much. Very very much as a matter of fact.

OSINF is not just the Internet[1]. It is also printed sources, or radio, or TV. It is also human sources, discussion groups, national archives, registries, etc. It is also handbooks, libraries, NNTP, FTP, radio, the Deep web. Almost all ’information’ out there is OSINF. Do not make the mistake of thinking that OSINT is limited to just the www. It is everything.

One of the largest libraries in the world, the Library of Congress, holds more then 170.000.000 items. They collect (almost) everything, from books, maps, CD’s, records, journals, newspapers to about anything. The British Library, the Vatican Library, the Library of St. Petersburg, all the same thing. Massive amounts of information all systematically organised and validated.

Commercial information providers such as Lexis-Nexis, Proquest Dialog and Factiva, hold thousands and thousands of databases full text online. Lexis-Nexis for example, currently offers more then 38.000 databases with an estimated total number of documents exceeding 8.500.000.000. The Internet as an information source. Never ever ask an OSINTian to ”find everything about...”. I am serious here. Internet search engines such as Google hold roughly 8.000.000.000 records, allbeit unvalidated and untested so there is a lot of slack in there.

The Internet is only a small part of what is available. Thinking that all information is available in electronic format and indexed by an Internet search engine is an assumption. And a wrong one too. The Internet has a few other ’sources’ of sometime valuable information, such as FTP, NNTP, IRC, POP and SMTP. IRC is still being used, for instance by child traffickers. But also by computer programmers and many others. NNTP holds tens of thousands discussion groups used until today, about any conceivable subject. The binary

[1] Technically, the Internet is a network of computers all running the TCP/IP protocol. In this paper we use the phrase Internet mainly for information sources found on the Internet

part is mostly the illegal part were stolen books, software, music etc is offered and asked for. POP is not just about reading e-mail. Thousands of ListServ discussion groups make use of of POP and SMTP to manage discussions.

The HTTP is but one service on the Internet, it is large, there is a lot, but it is only one.

And Google?

But all we need is Google because they have almost everything is it not?

Lets consider the size in number of documents of Google. We have seen already that the world of open source information is very much more then just the Internet. First of all, there is the pre-Internet world of printed information and information available online via commercial information providers. The 'Internet' is only part of that. The Internet consists of many different services as we have previously seen. The WWW is only one of them ().

The WWW consists of a wide array of web based services, a large part of which is in the so called deep web. The deep web is the part of the WWW that is not indexed by search engines. In other words, the part of the WWW that you will not be able to find (unless you know the exact URI).

The question then is, how large is the deep web with respect to the total www? According to Sherman and Price, the deep web is estimated to be 50-500 times larger than the WWW. This implies that only 2% or even just 0.2% of the WWW can be found using a search engine.

So how many Internet search engines are there and how large is the overlap? The answer to the first one is easy: less then 10[2]. Surprisingly, the overlap between the major search engines is just 20%, an exact explanation for this is unknown.

And Google?

Google is just one of these search engines. Now you realise how limited the Google search engine is with respect to the overall global information landscape? Google is large, very large, but in OSINF, as a information source, it plays a minor role.

Thinking that the answer you are looking for is available somewhere on the Internet is an assumption. Thinking that within that Internet domain Google has the answer is the second assumption. Both are wrong.

How to do OSINT

The osint Intelligence Cycle is a perfect explanation of how an osint research process is run. In short, there are three stages in a osint process (). The system was developed since 1990

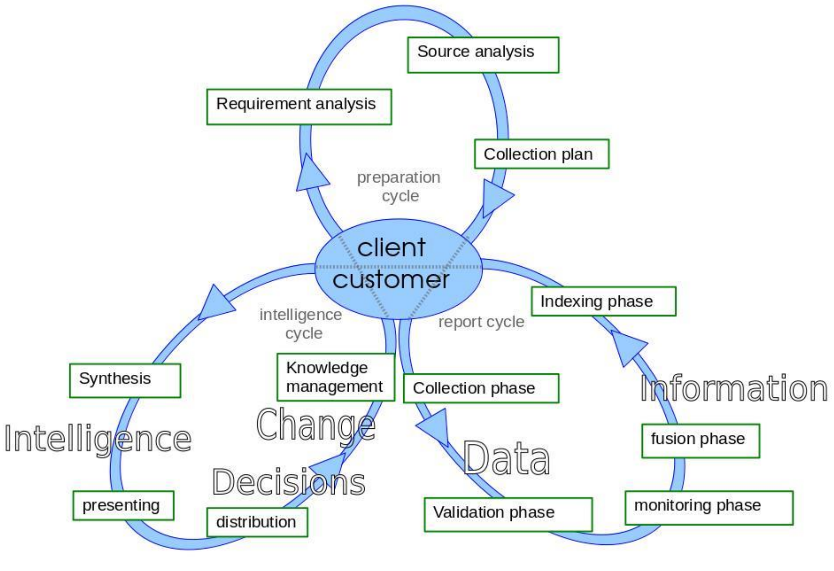
[2] Considering only the major international search engines

when starting to create the OSINT Branch for the Dutch Defence Intelligence & Security Service. The system since then became known as Arno's OSINT Methodology.

The cycle consists of three parts or wings, with the customer in the middle. In many existing intelligence cycles, the customer is not even mentioned and plays no role in the process. In our opinion, any intelligence production process is pointless if there is no match with a customer. The customer is the pivotal point of all OSINT operations.

The first wing is the exact information requirement of the customer. It's three parts make sure that the osint researcher knows exactly what the precise requirements of the customer are. Since the requirements can change quickly over time, the loop back makes sure that the outcome of the first wing is matched with the current requirements. If there is a match, we can continue to the second wing. If not, we shall do the first wing again.

The second wing is the search and collection phase. It is about creating a workable collection plan, validating the retrieved information on certain quality and reliability criteria, monitoring developments, a fusion phase were all validated information from a large variety of sources is brought together and deduplicated. Finally, the information is indexed with meta data to make it internally retrievable. Then we go back to the customer again to match the results with the requirements. If the customer is happy we can stop at this point. If not, we can do the second wing again. If the customer likes to have an analysis of the results, we continue with the third wing.



data – information – intelligence – decisions - change

© Reuser's Information Services 2017. RIS Intel Cycle v3

OSINT Intelligence Cycle

The third wing is the analysis and distribution wing. Here, the open source information results are matched with information from other, closed or covert, sources and analysed. We call this step synthesis, since analysis is done at every step. Presenting and distribution of intelligence are obvious challenges on their own. A fantastic OSINT report with a great analysis can be completely destroyed by a poor presentation (either orally or in writing). And distribution can be a challenge too, since in the intelligence community there still is up to today a strong tendency to write fantastic intelligence reports, but then not publish them, because the content is secret. The last step is knowledge management, recording the lessons learned, the methodologies used, the know how, in such a way that the knowledge obtained can be used again the next time.

Cases of OSINT

Let's have a look at two examples of actually doing OSINT.

The first is on trying to find out if Iran is actively working on the development of chemical weapons. It is a wonderful example of a combination of open sources, free as well as fee-based OSINF with advanced search techniques. This example demonstrates that OSINT is about knowing sources, that OSINT is about knowing online databases, that OSINT is about advanced search techniques. It also shows the role of internet search engines such as Google. Google is not used to find the answers to questions, Google cs is used to find authoritative sources.

The second one is a little bit more technical. It shows how to check a webpage for validity finding out that the person does not exist and the information on that page is fake. But let's start with Iran first.

Iran and chemical weapons

The assignment is clear: we like to find out if Iran is importing raw materials for the production of chemical weapons[3]. How to proceed using OSINT methodology?

The first step is, find out what exactly is a chemical weapon? Sources to be used here are the glsopcw to look up a general description of chemical weapons. We can also use the glsfas to do that, or any other handbook or organisation like the World Health Organisation, the US National Library of Medicine, the Center for Disease control and Prevention, or even a good old newspaper like the New York Times. What we learn from this is, that ordinary pesticides can be used as chemical weapons.

The second step is, how many pesticides are there? What are their names? Are their any synonyms for pesticides? We can use a variety of OSINF here. Either use handbooks, encyclopedias or some authoritative source. We can also make use of fee-based OSINF, like Proquest Dialog for instance. We can use one (or more) of their databases for chemical substances like Derwent Chemistry Resource, or Beilstein facts, ChemSearch to name a few. Also consider Pesticide Factfile. The query looks somewhat like the following:

[3] This example was first presented by the author at a NATO OSINT Meeting in Washington D.C. in 1995. The example is edited a little since then.

```
? B 390, 398, 355, 306 ? S PESTICIDES/na,de ? MAP SY T S1 ? SAVE TEMP
```

All these databases allow searching on nicknames, synonyms etc. to return a list of alternative names of all kinds of pesticides, insecticides and the like. On line 1 we start the appropriate databases, in line 2 is the query to search for the phrase pesticides in the name field and in the descriptor field, in line 3 we save all the synonyms temporarily in set 1, and in the final line we save this set temporarily on the Dialog server.

Now that we have a list of common nicknames, synonyms etc. for pesticides, we proceed with database number 571 Piers Exports. Let's do a search in this database to find out how many pesticides were exported from US harbours to Iran by launching our earlier saved query, sorting the records and creating a online report from all this. The queries look like:

```
? B 571 ? EXS ? SORT S1/ALL/CN,LB,CO ? REPORT S2/ALL/CN,LB,CO
```

This will produce a report that shows how many pesticides were exported to what country, with export date and manufacturers name.

```
Country of
Non-U.S.    Weight                                         Date
Port        (Pounds)  U.S.-based Company                   Shipped
----------  --------  -----------------------------        -------


ARGENT         5,388  BAYER                                031214
ARGENT         7,771  NA                                   041226
ARGENT         8,049  E I DUPONT DE NEMOURS                040527
ARGENT         8,049  E I DUPONT DE NEMOURS                040527
ARGENT         9,720  BAYER                                040718
ARGENT        11,661  NA                                   050407
ARGENT        15,878  NA                                   041126
ARGENT        17,641  NA                                   041111
ARGENT        17,642  NA                                   050506
ARGENT        17,648  BAYER                                040509
ARGENT        20,873  NA                                   050216
ARGENT        31,178  BAYER                                031214
ARGENT        31,755  NA                                   041211
ARGENT        31,812  KEY INTL SHPG                        040526
ARGENT        33,690  KEY INTL SHPG                        050409
ARGENT        34,593  KEY INTL SHPG                        050109
ARGENT        35,228  KEY INTL SHPG                        041018
```

So if this first rough estimate of the number of pesticides exported to Iran is now known, lets move on to step four. How much pesticides are within reason needed - on average - per hectare agricultural land? Again, let's find a authoritative source for this one. Such as the European Environmental Agency which gives a total pesticide consumption per hectare agricultural land.

If the rough estimate of required amount of pesticides is now known, the next question is how much agricultiral land does Iran actually have? In comes the good old CIA World Fact Book, to learn that Iran has about 9.78% arable land.

Finally, what follows is a little computation. We have 1) the amount of import of pesticides into Iran, t2) the number of pesticides needes per hectare of agricultural land, and 3) the amount of agricultural land in Iran. What follows is a very simple calculation: `1 - (2 * 3)` to find that Iran is importing way to much pesticides to within reason be used for agricultural purposes.

The uninitiated may start to think that this looks like a poor analysis. It is not. It has nothing to do with analysis. The example only shows part of a solution to find authoritative information. That's it.

The case of Mr. Bervoets

How to make money fast via the Internet? Simple, just browse the website of Mr. Frederik Bervoets who explains how he as a lorry driver with no more than secondary school discovered a method to get really rich very quickly. He created a manual on how to do that and made it public, available for everybody. It is really easy to get rich he explains, there is no startup costs and he does not sell any products. There is also a lovely picture of the man having a drink on a beach.

The question is, is this true? Is the website real or could this website be a scam? Using OSINT techniques it quickly becomes clear what is going on.

The first thing we do is a reverse image search on the picture. Reverse Image Searching is a technique supported by most Internet search engines to search on graphics to find websites with the same or similar picture. The technique is useful to validate the reliability of images, to find the origin of images and possible manipulation of images.

The result is most interesting. Google returns a bunch of websites with exactly the same picture and the same text, but in a different language and with a different name for the author (see . The man is called Scott Evans, or, Edgar Morgan, or, Thomas Stodola. He is also called Andy somewhere. The text on all these websites is almost identical, except that the age of the author differs.

Another clue that may be an indication that this website is a scam is checking the domain name. The domain name is `frederikbervoets.com`. The top-level domain is which (back then) draws attention since the website is in Dutch and aimed at The Netherlands. Why is the top-level domain not simply ? Lets do a WHOIS domain name to find the owner or registrant of this website to get a clue. The domain record shows that it is registered via Domains by Proxy, LLC, a subsidiary of GoDaddy.com, one of the largest providers in the world. Domains by Proxy is a company that will register domain names without mentioning the identity of the owner. That is interesting. Why would Frederik Bervoets do that? What is there to hide?

Let's use another tool to find out more. Traceroute will show all the nodes between our computer and the target website (unless firewalls pop up). Just take the final IP address and do a lookup on that address. In this case, we find an address Smallmead Road, in Reading, Berkshire, United Kingdom. With telephone number, e-mail address and more

Again, a simple example of using OSINT techniques to find out more about something. This example shows that OSINTians not just need a plan of action, they also need some technical skills too. Know about the organisation of the Internet, domain names, IP addresses and network topology. This is not cyber skills, that goes very much further.