Analysis of the 'Open Source Internet Research Tool': A Usage Perspective from UK Law Enforcement

Joseph Williams[1](✉) and Paul Stephens[2]

[1] School of Engineering and Computer Science,
University of Hertfordshire, Hatfield AL10 9AB, UK
j.williams30@herts.ac.uk
[2] School of Law, Criminal Justice and Policing, Canterbury Christ Church University,
Canterbury CT1 1QU, UK
paul.stephens@canterbury.ac.uk

Abstract. Internet Intelligence and Investigations (i3) are a fundamental investigative tool of the modern law enforcement official (LEO) in an always-connected online era. Ensuring LEOs follow good procedure for such investigations is critical for both law enforcement and society, as it ensures consistency, rigor and transparency.

Procedural issues lie with online evidential capture, however. For example, it is not feasible to directly apply digital evidence methodologies one would for 'offline' digital forensics; instead, one must apply best practices and a consistent approach. How those best practices and consistent approaches apply will typically fall to individual forces. One such tool in the arsenal of law enforcement is the 'Open Source Internet Research Tool' (OSIRT), a free all-in-one browser that assists law enforcement in conducting i3 in a standardized manner.

This paper analyses and discusses the results of 32 questionnaire responses from serving LEOs in the UK and their use of OSIRT. Results showed that LEOs found OSIRT to be helpful to them and compared to their previous method of conducting online investigations, OSIRT offered an improved system to conduct online investigations in many instances.

Keywords: Internet intelligence and investigations · OSINT · Digital evidential capture

1 Introduction

Law enforcement in the UK conduct Internet Intelligence and Investigations (i3), formerly Open Source Research, as part of their routine inquiries and roles. Previous work in the area of i3 showed that officers used a variety of different tools to conduct online investigations. Tools would often depend upon whatever software would be accessible, largely in relation to cost. This paper discusses Open Source Internet Research Tool (OSIRT), a free all-in-one web browser that was designed in collaboration with the

Published by Springer Nature Switzerland AG 2020
N. Clarke and S. Furnell (Eds.): HAISA 2020, IFIP AICT 593, pp. 341–352, 2020.
https://doi.org/10.1007/978-3-030-57404-8_26

UK's College of Policing to assist law enforcement of all skill-levels to conduct i3 in a rigorous and standardized manner. This paper continues the discourse from research [1, 2] conducted by the primary author and discusses the need for such a tool and evaluates and discusses 32 questionnaire responses from UK law enforcement that used OSIRT over a period of one month to two years.

2 Background

Knowing why law enforcement conduct investigations online may be obvious in an always-connected Internet-driven world, but matters arise surrounding the capture of such online artefacts. For example, these concerns can range from individual user skillset to legal and ethical issues. This section reviews and discusses prevalent issues UK law enforcement face when conducting i3 from a legal, ethical and practical perspective. The section then provides some background into OSIRT and its use in UK law enforcement.

2.1 Policing and Digital Crime in the United Kingdom

Her Majesty's Inspectorate of Constabulary and Fire & Rescue Services (HMICFRS) issued a report in 2018 outlining the importance of digital crime in policing. Data collection for the report took place over two months by visiting six police forces. The report uses examples of victim statements and how the police handled their reporting of the crime they had suffered.

The report stresses how integral technology is in modern society and how police must respond to the growing demand. The HMICFRS makes clear "[…] it is no longer appropriate, even if it ever were, for the police service to consider the investigation of digital crime to be the preserve of those with specialist knowledge." [3, p. 5]

HMICFRS stresses that all officers must understand handling and managing digital crime. The report from the offset sets out that regardless of job role, whether it is neighbourhood policing or anti-terrorism, officers must have the knowledge and skillset to police in the digital age, and that it is no longer a specialist's domain. The HMICFRS does acknowledge that to achieve the digital skillset required, those officers require to be trained in the technology they are meant to investigate. HMICFRS describe a "mixed picture" [3, p. 12] of officers' understanding surrounding digital crime, particularly highlighted by a response from an officer, "I am 46 years old. I do not have a computer; what do I know about Facebook?" [3; p. 30].

2.2 Legal, Procedural and Ethical Issues

Previous research by the author [2] has shown that law enforcement officials face several issues around online evidential capture. This section seeks to provide a background of the current legal, procedural and ethical issues of i3 and the prevalent statutes surrounding i3 in the context of 'open source' (i.e., publicly available) information.

Regulation of Investigatory Powers Act 2000 (RIPA) [4] is a key piece of legislation to look at when LEOs need conduct i3. However, given that RIPA pre-dates the modern era of social media platforms (e.g. Facebook was founded in 2004, Twitter in 2006) it

largely covers covert interception of communications from technology available at the time. Communications like email, SMS messages and telephones all comfortably fall under RIPA's authority but, unsurprisingly, it does not mention anything about social media.

However, RIPA's usage for i3 is necessitated upon the 'level' in which the investigation is being conducted. The levels, in order of 'unlikely to require RIPA authorization' to 'almost certainly requiring RIPA authorization', are: overt, covert 'core', covert 'advanced', network investigation and 'undercover'. For example, a publicly available 'open source' investigation at Level 1 is unlikely to require authorization under RIPA as the investigation makes use of searches using a search engine. The reason for using covert techniques, particularly at levels 2 and 3, is to minimize the 'footprint' of the investigating officer; i.e., the digital trace left behind when visiting a website. For example, an IP address, Internet Service Provider and location could show a law enforcement official from a police computer was visiting a website. "Covert" at these levels of open source capture focuses more on protection for the officer, and police network, as a counterintelligence and countersurveillance measure. Plainly, the higher the level the more training is required. For example, level 1 usually only requires basic training around force policy of computer usage.

Despite no legislation like RIPA to provide concrete structure, and regardless of the 'open' nature of i3, LEOs must still follow procedures and guidelines. The Association of Chief Police Officers (ACPO[1]) in the *Online Research and Investigation* manual lays out one such set of procedures. The 'Guiding Principles' state that viewing open source information "does not amount to obtaining private information because that information is publicly available" [4], and due to this it is "unlikely to require authorization under RIPA" [5]. However, ACPO [4] note that while the open sources may be collected, it must be "necessary and proportionate" and "does not necessarily mean a person has no expectation of privacy" [4]. Expectations of privacy are set out under Article 8, a right to respect for private and family life, under the European Convention on Human Rights (ECHR). Under the Human Rights Act (1998) [6], decisions when handling personal information must be "necessary" and "proportionate". Kent Police use the JAPAN test when handling personal information [7]. While the JAPAN test itself may not be followed by all police forces, its concepts will be. For example, authorization, necessity and proportionality are the backbone of UK policing, and form part of statue laws such as RIPA. Additionally, auditability and justification is guided by NPCC principles, along with data protection laws.

Case law itself provides few guidelines for the digital investigator when conducting open source research. A notable case is Bucknor v R [2010] EWCA Crim 1152, in which Bucknor appealed against his conviction of murder. The judge ruled in the initial case that evidence presented from the social networking sites Bebo and YouTube were admissible. While the initial conviction was upheld, the judgement from the appeal means any evidence taken from the Internet must have full provenance. That is, when (the date and time) and where (the website) the evidential artefact was obtained should be audited.

[1] Now the National Police Chiefs Council (NPCC), but the previous ACPO and its guidelines still greatly impact force policies. ACPO principles are still widely used and trained.

The General Data Protection Regulation (GDPR) came in effect in May 2018 and has had an impact on how law enforcement within the UK and the European Union (EU) manage personal data. The GDPR provides citizens (termed "data subjects") with greater control over their personal data from "controllers" (i.e. those who control the data subject's personal data). Data subjects now, trivially, can access and remove personal data upon request.

GDPR provides member states of the EU provisions on how to apply GDPR, and in the UK this brought in the Data Protection Act 2018, superseding the 1998 Act of the same name. The Data Protection Act (2018) [8] covers aspects that "fall out of scope of EU law" [9], such as national security and how "intelligence services" manage personal data; this is covered by Part 4 of the Act. However, Part 3 of the Data Protection Act (2018) covers "Law Enforcement Processing" and provides six "protection principles" in Chapter 2 of the Act for those managing personal data for law enforcement purposes.

Law enforcement are afforded exemptions from the Data Protection Act (2018) but must follow 'protection principles' within the Act as there are themes of necessity and proportionality when handling sensitive data.

2.3 i3 Capture and OSIRT

From both a technical and procedural perspective of conducting i3, officers used a variety of software tools that would vary both in price and quality, with no standard toolset. To bring about standardisation, the College of Policing's Research, Identifying and Tracing the Electronic Suspect (RITES) course recommended several capturing and productivity tools. Trainers on the RITES course soon discovered the cognitive overload this had on the cohort, who would often spend more time learning to use the tools than learning about i3 techniques.

The problem highlighted above prompted the creation of Open Source Internet Research Tool (OSIRT); an all-in-one browser for conducting i3. OSIRT's creation followed the user-centred design (UCD) method, with a two phased development using the software engineering methodologies 'throwaway prototyping', for the prototype version, and 'incremental and iterative development' for the release version.

OSIRT has since been integrated into the RITES course, which trains over 100 officers a year, and provides a feedback outlet for OSIRT. Officers are also using OSIRT back on-the-job.

OSIRT's target audience is those officers, particularly case officers who require a streamlined method of conducting online investigations in terms of systematic evidential capture at Levels 1-3 (overt to covert). However, given OSIRT is a web browser it also has broader uses and can be used for Levels 4 and 5.

2.4 i3 and Open Source Intelligence-Style Browsers

For completeness, this section provides an overview and discussion of several popular OSINT-style 'OSIRT competitor' browsers and applications. Note, the latest version of Forensic Acquisition of Websites and Hunchly came out after OSIRT was released.

Oryon. OSINT Browser. Oryon OSINT Browser (Oryon) is a free browser built using Chromium, making its look and feel much like Google Chrome. The browser itself makes use of a plethora of add-ons and extensions, largely available via the Chrome web store, which makes Oryon extremely feature rich. While Oryon boasts more than 60 pre-installed add-ons, this leaves the interface brimming with icons to the point where it is bordering on overwhelming.

Oryon's overall design leaves the impression it is for those who are advanced computer users who can happily make use of and understand the needs of the add-ons. Oryon does not offer hashing capabilities for files, or report exporting.

Forensic Acquisition of Websites. (FAW - https://en.fawproject.com/) is not a browser designed for conducting open source investigation, but it is a browser designed for law enforcement purposes and reviewed for this reason. Initially, this review focused on the free, and only version, of FAW that was made available in November 2014 and not updated until early-2017.

The 2014 version of FAW was very much a simple, visual website saving application whereby a user visits the page to they wish to capture and clicks the "acquisition" button. FAW would then download the contents of the website and place it within a directory structure. All items acquired were date and time stamped and logged in an XML file. The browser did not offer anything beyond this capturing ability in this version.

FAW lay dormant for several years but came back with an updated version in 2017 that replaced the main browser with CefSharp. While FAW was initially a free product, a tiered pricing model was adopted from FAW version 5. This saw a free, professional and law enforcement licences added. The paid for versions unlock, amongst other features, Tor and user-agent spoofing.

Hunchly. Hunchly (https://www.hunch.ly/) is a paid for extension for the Google Chrome browser. Hunchly costs $129.99USD a year for a single license, or $349.99USD per three licences with a 20% saving for more than three users. This review is based on the major update version of Hunchly released in April 2018.

Hunchly sits within the Google Chrome browser and automatically logs webpages when a user visits by placing them within a local case file; this is the big selling point of Hunchly. Case files can then be accessed by means of the "dashboard", a separate application outside of the browser extension. Additionally, Hunchly contains features such as file attachments, automatic hashing, social media ID extraction and report exporting to both docx and PDF. Hunchly is a very capable addition to the OSINT browsing family, plus has the benefit of being cross-platform because it is a browser add-on.

However, there are several issues with using Hunchly that may impact its use, both from a legal and ethical perspective. In particular, the automated saving of every webpage visited creates an interesting dilemma. The immediate question: is it fair for law enforcement to make automated and automatic copies of webpages they visit without the need to make a conscious decision to do so? Previously, it was shown saving data using an automated means is a breach of Facebook's terms and conditions, but there are ramifications further afield than just a website's policy.

The process of "do first, ask questions later" is, in the opinion of the author, the wrong approach; particularly surrounding law enforcement's collection of personal data.

This chapter has shown that law enforcement need to take a careful and considered approach; one that focuses of necessity and proportionality. Is it then necessary and proportionate to automatically store carbon copies of all websites visited, without any interaction or acknowledgement from the investigating officer? The Data Protection Act (2018) explicitly states that personal data collection must be "Adequate, relevant and not excessive", and debatably, visiting a webpage may be "relevant" to the investigation but arguably that maintaining a copy of every webpage is excessive, particularly with only having to optionally justify that capture with a note. Of course, users can simply delete these traces if not required, but then the audit trail is lost.

3 Method

This paper looks at questionnaire data collected from officer's usage of OSIRT back on-the-job, how they were trained to use OSIRT and their thoughts and feelings.

The questionnaires were distributed to officers via the Police Online Knowledge Area (POLKA)[2] and were completed by 32 participants.

Questionnaires are an indirect method of data collection and are a traditional, efficient method of data collection, as the researcher is not required to be present during their administration. Questionnaires can obtain both quantitative and qualitative data, depending upon the type of questions asked (i.e. open or closed). Questions can generate diverse opinions from respondents, which can then lead to generalisability of any conclusions derived from the responses. Responses are gathered in a more standardised way, particularly when compared to interviews [10].

Limitations surrounding questionnaires are the potential for non-response, particularly for self-administered questionnaires, the consequence of a low/non-response rate may effective generalisability of the results. Additional limitations are that respondents may embellish their answers in order to provide a 'socially acceptable' response; this is known as social desirability bias [11].

3.1 Sample

This section details the officers who participated in the questionnaire and provided details about themselves (Table 1). These responses were optional, so results may not add-up to 32.

As expected, there is a mix of job roles and experience. There is a high proportion of analysts and detective constables, which is not surprising given that OSIRT is a hands-on tool designed specifically for investigators.

4 Results and Discussion

4.1 Previous Tool Usage

These results are typical from what has been previously discovered [1]. Popular tools such as Microsoft Excel and Word would be used to maintain the audit log, and various

[2] POLKA closed in January 2020 and was since replaced with 'The Knowledge Hub'.

Table 1. Participant roles and average years active per role.

Role	n	Years active (avg)
Trainer	1	10
Police officer	6	12
Intelligence researcher/analyst	9	3
Detective sergeant	2	17
Detective constable	10	11
Digital media investigator	3	5

other tools and add-ons to capture. In this questionnaire, the browser extension *Fireshot* was the most popular screenshot tool. Even with a pool of 32 responses, it shows the disparate use of different tools that OSIRT has ultimately went on to replace (Table 2).

Table 2. Breakdown of previous tool usage.

Productivity tool	n
Excel/spreadsheet	9
Unspecific add-ons/extensions for browsers	6
Word	5
Fireshot	5
Karen's Hasher	4
Notepad/(++)	3
Camtasia/screen recording	3
Whois? Add-ons	3
Snagit	2
None	2
Ashampoo	1
One note	1
Windows screenshot	1
HTTRACK	1
Tor	1

4.2 OSIRT Usage

Table 3 breaks down the how long the participants have been using OSIRT. Of the respondents, 63% have been using OSIRT for a year or more. Most respondents, 80%, have been using OSIRT for at least 10 months. Those users who have been using OSIRT for two or more years are likely to be users of the prototype and have been using OSIRT around its initial release.

Table 3. How long participants have used OSIRT

How long using OSIRT	n
Over two years	11
Over one - two years	8
four months to one year	9
one to three months	1
Less than a month	2

The interesting aspects of the results of average weekly usage in Table 4 show there are two groups of users. One that uses OSIRT for a not insignificant amount of their work, 11 use OSIRT between 11 and 25 h a week, and those that use it in a more casual manner; 17 use it for two hours or less a week on average. These are not particularly striking results, as not all officers will be tasked with conducting open source research all the time. Some respondents are likely to be "satellite" open source researchers, in that they may start an open source investigation for the dedicated team to start later. For example, starting a case during the night for a Digital Media Investigator to pick up in the morning.

Table 4. The average weekly usage in hours

Average weekly usage	n
Over 25 h	0
16-25 h	2
11-15 h	9
7-10 h	1
3-6 h	1
1-2 h	12
Less than 1 h	5

4.3 How Officers Were Trained to Use OSIRT

While OSIRT is utilised during the RITES course, it is often trained as part of in-house training packages, as seen in Table 5. 25 respondents were either trained directly as part of an internal training package, or by a colleague. Unsurprisingly, internal training is popular as it is cheaper than sending officers to training sessions. Sending officers away will mean losing a resource for a week on top of the cost of the training itself. Additionally, keeping training in-house means officers can be trained to that force's operating procedures and standards. While the RITES course teaches open source research techniques, it can only discuss methods and procedures in a generic manner for the diverse cohort; ultimately this will boil down to force policy. It is not uncommon to see officers attending the RITES course in order to then feedback and train in-house.

Table 5. How respondents were trained to use OSIRT

Trained to use OSIRT	n
Colleague	7
RITES course	2
Self-guided	4
In-house	19

4.4 Does OSIRT Capture All Relevant Data?

This free-form question, with responses in Table 6, offered the respondents a chance to provide feedback on whether OSIRT captures relevant data as part of their open source investigation. Word frequency analysis of the text showed there were 29 occurrences of the word 'yes'. Two respondents noted an issue surrounding video capture.

4.5 Has OSIRT Enhanced the Capability to Conduct Internet Investigations?

This free-form optional question generated 29 responses. Of the responses, 22 started their sentence with "yes" and a further 4 responses were positive in nature. One comment from an officer who has used OSIRT for over two years notes OSIRT's integrated tools and the fact it was designed specifically for law enforcement as a reason for why it has enhanced their capability:

"It has [enhanced my capability], but, it's the fact that this tools places all the relevant functionality of other tools all in one place that is specifically designed for Law Enforcement and the challenges that we face around continuity of evidence.
It also gives peace of mind as we know that all data is locally held and OSIRT is not reporting back to any servers, meaning we can trust it for security around our information."

Table 6. Freeform response for OSIRT's capture abilities

Response	n
"Yes" or "yes"	21
Yes, although the ability to download videos from more websites would be great	1
Yes, the tool is particularly useful for audit and reporting	1
Only current issue is video capture	1
Yes - I always video capture my screen and produce this in evidence	1
For me it does yes	1
Yes. I particularly like the screen recording options and the automatic page logging	1
I struggle capturing video and sound	1
Yes - extremely easy to use and professional means of recording what we do on open source	1
Yes - and more!	1
Yes and then some	1

Of the negative comments, those who said OSIRT has not enhanced their capability, still provided positive feedback "it has enhanced our methods of recording our research and auditing process" and "It has not enhanced - may be user error but it is great at tracking my movements evidentially".

Word frequency analysis showed "easier" was mentioned 6 times. In context, these comments all noted that OSIRT had made conducting Internet investigations easier, with one comment even mentioning it "made my job much easier".

The notion of professionalism OSIRT brings to respondents was also emphasised via word frequency analysis with three participants mentioning how OSIRT provided an output that is "more professional", with another respondent saying, "It has added professionalism to our [Internet investigations]".

It's all I ever knew… For one respondent, they had "only ever used OSIRT" to conduct their open source research. While this is only one respondent, it perhaps shows that for many incoming officers who are required to conduct research, OSIRT will be the de-facto piece of software they use. This will, speculatively, only increase as OSIRT has only been available for several years, so some of those officers who joined the force in 2016 will now be coming off probation into different roles, and perhaps require using OSIRT. This is also highlighted in the Sect. 4.1 (previous tool usage), where several respondents did not list tools as they had only used OSIRT.

4.6 Tool Usage Within OSIRT

Table 7 lists individual tool usage within OSIRT. The usage figures lend credence to the previous discussion during the analysis of SUS results surrounding the 80:20 rule. All tools within OSIRT are used, but of the 20 tools listed seven are used half of the time

with only four used at least two-thirds of the time. No individual tool is listed as 100% usage.

Table 7. Individual tool usage within OSIRT (total usage and total usage as a percentage)

Tools	n	%
Video screen capture	22	70.97
Audit log	22	70.97
Full screenshot capture	21	67.74
Snippet capture	21	67.74
Case notes	17	54.84
Report exporting	17	54.84
Tabbed browsing	16	51.61
Full webpage downloading	13	41.94
Timed screenshot	12	38.71
Saving page source code	12	38.71
Attachments	11	35.48
Video downloader	11	35.48
WhoIs? Finder	11	35.48
IP address saver	11	35.48
Facebook and Twitter ID finder	11	35.48
Extracting links on webpage	9	29.03
Tor (dark web browsing)	6	19.35
Exif viewer	6	19.35
Reverse image searching	6	19.35
History viewer	5	16.13

These figures certainly lend credence to Pareto's '80:20' principle as discussed previously. If we consider a tool to be 'popular' that is used by at least two-thirds of respondents, we see a ratio close to 70:30. Given the modest sample size, that is close to the original principle.

5 Conclusion and Future Work

This paper analysed and discussed the results of 32 questionnaire responses from UK law enforcement regarding OSIRT. Results showed that OSIRT greatly assisted officers in their Internet investigations, when compared to previous tool usage. The responses also highlighted the policing spectrum, where OSIRT was used by neighbourhood officers to Digital Media Investigators. For policing to continually be responding to the challenge

presented by digital technology they must adapt to their ever-changing surroundings, as the report by the HMICFRS commented.

Future work will look at distributing a similar questionnaire to a broader number of LEOs, as those that chose to fill out questionnaires were, arguably, 'fans' and users of OSIRT. This means that feedback focussed more on positive feedback from OSIRT fans. Access to those who do not use OSIRT, or do not like to use OSIRT, are harder to find because they are unlikely to reach out, or do not visit locations where OSIRT is discussed (e.g. police knowledge exchange forums).

References

1. Williams, J.: Creating and integrating a FLOSS product into UK law enforcement. In: Stamelos, I., Gonzalez-Barahoña, J.M., Varlamis, I., Anagnostopoulos, D. (eds.) OSS 2018. IAICT, vol. 525, pp. 117–127. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92375-8_10
2. Williams, J.: Legal and ethical issues surrounding open source research for law enforcement purposes. In: Skarzauskiene, A., Gudeliene, N. (eds.) Proceedings of the 4th European Conference on Social Media, Mykolas Romeris University. Vilnius, Lithuania. 3–4 July, 2017. ISBN 9781911218463
3. HMIC: Real lives, real crimes: A study of digital crime and policing (2018). https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/real-lives-real-crimes-a-study-of-digital-crime-and-policing.pdf
4. Regulation of Investigatory Powers Act (2000). http://www.legislation.gov.uk/ukpga/2000/23/contents. Accessed 25 Apr 2020
5. Association of Chief Police Officers: Online Research and Investigation, 16 September 2013. http://library.college.police.uk/docs/appref/online-research-and-investigation-guidance.pdf. Accessed 25 Apr 2020
6. Human Rights Act 1998. http://www.legislation.gov.uk/ukpga/1998/42/contents. Accessed 25 Apr 2020
7. Kent County Council: The JAPAN Test. https://www.kelsi.org.uk/__data/assets/pdf_file/0003/26706/Japan-Test.pdf. Accessed 25 Apr 2020
8. Data Protection Act 2018. http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted. Accessed 25 Apr 2020
9. Information Commisioners Office: Data Protection Act 2018 (2018). https://ico.org.uk/for-organisations/data-protection-act-2018/. Accessed 25 Apr 2020
10. Milne, J.: Centre for CBL in Land Use and Environmental Sciences, p. 1. Aberdeen University (2009)
11. Grimm, P.: Social desirability bias. In: Sheth, J., Malhotra, N. (eds.) Wiley International Encyclopedia of Marketing (2010). https://doi.org/10.1002/9781444316568.wiem02057