Università degli Studi di Camerino

FACOLTÀ DI SCIENZE E TECNOLOGIE
Corso di Laurea Triennale in Informatica

Il mondo dell'Osint

Nico Asciutti
Alessandro Fraticelli
Edoardo Iommi

Relatore:
Fausto Marcantoni

Anno Accademico 2021-2022

Indice

Sommario

Studio e approfondimento del mondo dell'intelligence open source, comprendenti prove e esecuzioni di strumenti che operano secondo questa filosofia.

1 Introduzione

Il valore delle informazioni open source (OSIF) per integrare l'intelligence riservata è stato riconosciuto da tempo, ma la crescente pervasività di Internet, l'ascesa dei social media e l'analisi dei big data negli ultimi due decenni hanno rivoluzionato l'*open source intelligence* (OSINT). Queste nuove fonti aperte hanno anche il potere di sostituire e/o integrare l'accesso a informazioni che un tempo potevano essere ottenute solo attraverso strumenti e metodi di intelligence tradizionali più pericolosi e costosi.

Sebbene l'OSINT esista da più di 50 anni, la sua definizione e le sue caratteristiche come disciplina di intelligence sono ancora oggetto di dibattito. In un documento del 2011 emesso dall'Office of the Director of National Intelligence[1], OSINT è stata definita come "Intelligence prodotta da informazioni pubblicamente disponibili che vengono raccolte, sfruttate e diffuse in modo tempestivo a un pubblico appropriato allo scopo di soddisfare uno specifico requisito di intelligence."[2]

Con l'avvento di internet e l'ascesa dei social media, l'OSINT sta diventando più complesso in termini di fonti e metodi. Singoli individui sono ora in grado di rendere disponibili informazioni in modi mai esistiti prima. Combinando la potenza di calcolo moderna con tecniche di data science è possibile registrare ed elaborare grandi quantità di dati pubblicamente disponibili.

Prima di questa rivoluzione, l'attività principale nell'OSINT era principalmente la traduzione: rendere accessibili notizie estere agli analisti di intelligence. Una volta tradotto un rapporto open-source in gran parte originale, gli analisti all-source potevano incorporarlo in un prodotto di intelligence finito. L'attività moderna di OSINT richiede spesso fasi di acquisizione, elaborazione e sfruttamento più complesse per produrre un prodotto open source che può quindi essere integrato in un prodotto finito all-source. Inoltre, le informazioni spesso provengono da individui, il che introduce nuove complessità nella protezione della privacy delle singole persone. Tutte queste modifiche richiedono una definizione più solida di OSINT, poiché i dati open source possono assumere molte forme.

In considerazione della natura mutevole delle informazioni pubblicamente disponibili, si considera[3] il periodo attuale come la seconda generazione di OSINT. L'ascesa del personal computer negli anni '90, - periodo in cui è stato coniato l'acronimo OSINT - ha avuto un enorme impatto. Eventi come l'Onda verde iraniana nel 2009[2] hanno fornito un vivido esempio di come l'utilizzo di nuove forme di social media può fornire grandi moli di informazioni in tempo reale in un ambiente inaccessibile.[4] Si può datare il passaggio all'OSINT di seconda generazione al 2005, anno in cui l'IC ha creato l'Open Source Center,[5] in concomitanza con i cambiamenti che internet stava subendo in quel periodo, con la maggior parte dei contenuti online che si spostava su pagine Web dinamiche, contenuti generati dagli utenti e social media. Questa transizione è spesso descritta come l'emergere del Web 2.0.[6]

[1] Il *Director of National Intelligence* (DNI) è un alto funzionario del governo degli Stati Uniti che funge da capo della *US Intelligence Community* (IC), supervisionando e dirigendo l'attuazione del *National Intelligence Program* (NIP). Il DNI funge anche da principale consigliere del Presidente degli Stati Uniti d'America, del National Security Council e del Homeland Security Council per questioni di intelligence relative alla sicurezza nazionale.

L'*Office of the Director of National Intelligence* (ODNI) è composto da ufficiali di tutto l'IC.[1]

[2] `https://www.treccani.it/enciclopedia/onda-verde/`

2 Storia dell'Osint

2.1 Prima generazione di OSINT

La storia dell'utilizzo delle informazioni aperte risale all'emergere dell'intelligence come strumento a sostegno delle decisioni e delle azioni di un governo. Tuttavia, non si è mai avuto un approccio metodico fino a quando gli Stati Uniti non hanno aperto la strada all'istituzionalizzazione e alla professionalizzazione del monitoraggio dei media stranieri.

Una delle figure più importanti nella nascita dell'intelligence militare americana e, di conseguenza, dell'intelligence open source è senza dubbio William Donovan. Nato a Buffalo nel 1883, è cresciuto in una famiglia operaia di origini irlandesi, eccellendo a scuola e nel mondo accademico.

Dopo aver studiato giurisprudenza alla Columbia Law School e aver combattuto nella prima guerra mondiale, Donovan ha avuto una carriera di successo come avvocato internazionale. Durante tutto il periodo tra le due guerre, Donovan ha viaggiato per il mondo come avvocato, incontrando influenti personalità straniere e successivamente scrivendo rapporti su di esse per il governo degli Stati Uniti in veste semi-ufficiale.

Nel frattempo, con lo scoppio e l'inasprirsi della seconda guerra mondiale in Europa, il 26 febbraio 1941, il presidente americano Roosevelt ordinò che fossero stanziati 150.000 dollari per la creazione del Foreign Broadcast Monitoring Service (FBMS) sotto l'autorità della Federal Communications Commission. Il compito dell'FBMS era quello di registrare, tradurre, trascrivere e analizzare i programmi radiofonici di propaganda a onde corte che venivano trasmessi negli Stati Uniti dalle potenze dell'Asse.[7][8]

Donovan fece quindi pressioni sull'allora Presidente degli Stati Uniti ed ex compagno di studi nella Columbia University, Franklin D. Roosevelt per formalizzare il suo lavoro ufficioso per il governo degli Stati Uniti e l'11 luglio 1941, Roosevelt ha creato per Donovan la carica di "Coordinator of Information". In un primo momento, Donovan ha riferito direttamente al presidente. Prima di allora lo spionaggio era malvisto da molte persone nell'establishment della politica estera: nel 1941, mentre Hitler invadeva l'Europa e minacciava la Gran Bretagna, il Dipartimento di Stato aveva diciotto persone che lavoravano nell'intelligence. Dopo Pearl Harbor, la necessità di intelligence era chiara e il dipartimento di Donovan fu ribattezzato Office of Strategic Services (OSS), il precursore della CIA.

L'OSS aveva un intero ramo dedicato all'intelligence open source. Il ramo di ricerca e analisi dell'OSS ha raccolto meticolosamente dozzine di giornali e rapporti su trasmissioni radiofoniche da tutto il mondo, alla ricerca di foto e articoli che potessero rivelare informazioni cruciali sul nemico. Nelle parole di Donovan: "Anche una stampa irreggimentata tradirà ripetutamente gli interessi della propria nazione agli occhi di un osservatore scrupoloso".[9]

L'OSS ha monitorato le sezioni dei necrologi sui giornali regionali tedeschi, in cerca di notizie su importanti figure naziste. Le immagini di nuove corazzate, crateri di bombe e aerei sono state accuratamente raccolte e, una volta valutate, hanno permesso all'OSS di valutare lo stato degli eserciti dell'Asse. È da notare come le attività dell'epoca portate avanti dall'OSS ai suoi albori siano simili alle moderne indagini OSINT, anche se senza computer. Tra l'OSS e il suo corrispettivo britannico, lo Special Operations Executive, è possibile sostenere che le origini dell'intelligence open source risalgono a quasi un secolo fa.

Nel 1939, il governo britannico operò in maniera simile chiedendo alla British Broadcasting Corporation (BBC) di lanciare un servizio civile, e in seguito commerciale, di controllo della stampa giornalistica e delle trasmissioni radiofoniche straniere con il Digest of Foreign Broadcasts, in seguito intitolato Summary of World Broadcasts (SWB) e oggi noto come BBC Monitoring.[10] Come si legge dall'edizione del 1940 del BBC Handbook, l'obiettivo era quello di erigere una "moderna Torre di Babele, dove, con una concentrazione esemplare, si ascoltano le voci di amici e nemici".[11]

Nel frattempo, il 26 luglio 1942, il FBMS è stato ribattezzato Federal Broadcast Information Service (FBIS), in parte per farlo sembrare più simile a un'agenzia di guerra. A testimonianza dell'importanza dell'operato open source dell'FBIS, Charles B. Fah, allora capo ad interim della Sezione Estremo Oriente dell'OSS, scrisse

il 13 agosto 1942 al Direttore dell'FBIS Robert D. Leigh che: "La Sezione dell'Estremo Oriente dell'Office of Strategic Services ha trovato i vari rapporti della Foreign Broadcast Intelligence Service della Federal Communications Commission indispensabili nel nostro lavoro. Le trasmissioni giapponesi monitorate offrono la più ampia fonte disponibile di informazioni sugli sviluppi dall'8 dicembre 1941 in Giappone e nei territori da esso occupato. Le trasmissioni cinesi monitorate sono importanti indicazioni del pensiero e della morale nella Cina libera". Fah ha conclude chiedendo una maggiore copertura delle trasmissioni giapponesi. Nello stesso anno, il direttore R&A dell'OSS William D. Langer scrisse a Leigh che: "Senza il servizio di monitoraggio della Federal Communications Commission, la nostra conoscenza degli eventi attuali in Giappone sarebbe esigua e il fatto che vorremmo più di una cosa buona non implica una mancanza di apprezzamento di ciò che stiamo ricevendo. Posso parlare solo per noi, ma sono certo che ci sono altre agenzie a Washington i cui membri la pensano come noi. Washington i cui membri la pensano come noi».[12]

Quando fu sull'orlo di essere chiuso alla fine della seconda guerra mondiale, il FBIS fu rilevata dal Dipartimento della Guerra il 1 gennaio 1946 e trasferita alla neonata CIA (che di fatto sostituì l'OSS dopo il suo scioglimento nel settembre 1945) un anno dopo ai sensi del National Security Act del 1947, quando era già un'organizzazione "quasi matura, addestrata e disciplinata" dall'esperienza bellica. Vent'anni dopo, la storia ufficiale dell'FBIS della CIA ne descriveva "l'organizzazione e le responsabilità fondamentali [...] funzionamento e metodi di base" come sostanzialmente invariati.[8]

Dalla creazione dell'FBIS fino agli anni '90, l'ambito dell'analisi open source all'interno dell'IC americana è stato principalmente il monitoraggio e la traduzione delle fonti della stampa estera. Ci sono alcune differenze importanti tra questa versione storica di OSINT, la prima generazione, e la seconda generazione odierna. La raccolta di materiale è stata di particolare importanza nelle operazioni OSINT di prima generazione. L'FBIS gestiva 20 uffici in tutto il mondo per riuscire a raccogliere fisicamente materiale da utilizzare. Le ambasciate estere hanno anche fornito una piattaforma per la raccolta di materiale, oltre agli ufficiali diplomatici, gli addetti alla difesa agivano da raccoglitori di informazioni. Il Defense Attaché System è stato consolidato nel 1964-1965 sotto le autorità della Defense Intelligence Agency (DIA).[13] Alcune attività sono state nel tempo ridotte per concentrarsi su materiale ad alta priorità, ma il requisito principale per elaborare questo materiale era principalmente la traduzione; tuttavia, va notato che l'FBIS ha svolto alcune funzioni analitiche, principalmente l'analisi delle tendenze, sin dalla sua fondazione.[14]

Dopo la fine della Seconda guerra mondiale, i funzionari dell'intelligence esperti in open source continuarono ad aiutare analisti e funzionari nel complesso contesto della Guerra fredda. Ad esempio, gli analisti dell'FBIS e della Foreign Document Division (FDD) guidarono la CIA nell'individuare gli sviluppi dell'allontanamento tra Mosca e Pechino. I funzionari dell'FBIS e dell'FDD cominciarono a scorgere i segni della spaccatura sino-sovietica leggendo il materiale di propaganda all'inizio degli anni Cinquanta. Al contrario, alcuni ufficiali della CIA che operavano sotto copertura sbagliarono, insieme a molti altri osservatori, nel liquidare come disinformazione le prove open source per tutto il decennio successivo.[15]

Per tutta la durata della Guerra Fredda, infatti, l'OSINT diventò progressivamente una parte sempre più importante di tutta l'intelligence sull'Unione Sovietica, sulla Cina e su altri avversari. Negli ultimi anni della Seconda guerra mondiale, si ricercavano informazioni sulle capacità tecniche sovietiche solamente nei documenti ufficiali tedeschi, giapponesi e russi recuperati in guerra. Alla fine degli anni '50, la CIA e l'Aeronautica avevano scoperto una "miniera di informazioni" nel crescente flusso di libri e periodici provenienti dall'Unione Sovietica.[16] All'inizio degli anni '60, un informatore scrisse che "in totale, le fonti aperte forniscono probabilmente la maggior parte di tutte le informazioni utilizzate nella produzione di intelligence militare sull'Unione Sovietica".[17] Alla fine dello stesso decennio, un altro informatore scriveva della "marea di carta stampata pubblicamente" che sosteneva e minacciava di "sommergere" la Intelligence Community. Ha anche offerto un esempio del valore dell'OSINT: "L'intenso controllo della stampa e della radio nordvietnamite è stato un elemento di intelligence essenziale a sostegno dello sforzo [degli] Stati Uniti" nel conflitto in Indocina.[18]

Vale la pena notare che tutte le potenze hanno sfruttato l'OSINT durante la Seconda guerra mondiale e

la Guerra fredda per monitorare gli sviluppi esteri e per risparmiare tempo e denaro sui propri progetti. La pubblicazione aerospaziale statunitense Aviation Week, soprannominata "Aviation Leak" per i suoi scoop, era tra i periodici tecnici statunitensi che l'intelligence della Germania Est, tra gli altri, traduceva per monitorare gli ultimi sviluppi nel settore aerospaziale americano.[19]

Alla fine della Guerra Fredda, i progressi commerciali e tecnici avevano reso evidente il valore dell'OSINT. La radio, all'avanguardia negli anni Trenta, rimase una fonte fondamentale nella Seconda guerra mondiale e negli anni successivi. Quando i carri armati sovietici entrarono a Budapest nel 1956, ad esempio, i funzionari dell'intelligence di Washington si informarono sugli eventi attraverso i rapporti radiofonici. Un veterano della Directorate of Operations (DO) della CIA, riferendosi alla repressione della rivolta ungherese da parte di Mosca, scrisse: "È un fenomeno ben noto nel campo dell'intelligence che spesso arriva un momento in cui l'attività politica pubblica procede a un ritmo così rapido e fulminante che l'intelligence segreta, il lavoro degli agenti, viene superato dagli eventi riportati pubblicamente".[20]

L'attività dell'FBIS ha fornito informazioni di vitale importanza e spunti decisionali alla sfera militare americana durante il periodo della Guerra Fredda, tra cui i primi segni della rimozione dei missili sovietici da Cuba, l'avvisaglia del ritiro sovietico dall'Afghanistan e il contesto delle crisi in Ungheria e Cecoslovacchia. L'ottanta per cento delle informazioni utilizzate per monitorare il crollo dell'Unione Sovietica è stata attribuita a fonti aperte.[21]

La fine della Guerra Fredda ha comportato tagli al budget per la maggior parte delle istituzioni dell'Intelligence Community, ma ha creato una crisi in particolare per l'FBIS e le attività open source. Se da una parte il volume di OSIF stava aumentando notevolmente, dall'altra l'FBIS stava rapidamente perdendo risorse. Nel 1997, l'FBIS era a rischio di scioglimento come parte dei tagli al budget della CIA, ma è stato salvato da una campagna pubblica guidata dalla Federation of American Scientists.[22]

2.2 Seconda generazione di OSINT

Con l'avvento di internet e l'ascesa dei social media, il contesto dell'OSINT si è fatto più complesso in termini di fonti e metodi. Singoli individui sono ora in grado di rendere disponibili informazioni in modi mai esistiti prima. I leader dell'IC hanno riconosciuto che le sfide e le dinamiche del 21° secolo avrebbero portato più domanda di OSINT. Il vicedirettore dell'FBIS, J. Niles Riddel, al primo simposio internazionale sull'open source nel 1992, ha riconosciuto i cambiamenti nell'OSINT derivanti dall'ascesa del personal computer, dell'archiviazione digitale di grandi capacità, dei motori di ricerca e delle reti di comunicazione a banda larga. Credeva che tutti questi fattori avrebbero portato a una crescita esponenziale nella commercializzazione delle informazioni.[14] Nello stesso evento, l'allora vicedirettore della CIA, l'ammiraglio William Studeman, chiese "un cambiamento rivoluzionario nell'approccio dell'Intelligence Community alla gestione, alla raccolta, all'elaborazione e alla diffusione dell'open source."[21]

Prima di questa rivoluzione digitale, l'attività principale nell'OSINT era infatti principalmente la traduzione: rendere accessibili notizie estere agli analisti di intelligence. Una volta tradotto un rapporto open source in gran parte originale, gli analisti all-source potevano incorporarlo in un prodotto di intelligence finito. L'attività moderna di OSINT richiede invece spesso fasi di acquisizione, elaborazione e sfruttamento più complesse per produrre un prodotto open source che può quindi essere integrato in un prodotto finito all-source. Inoltre, le informazioni spesso provengono da individui, il che introduce nuove complessità nella protezione della privacy delle singole persone. Tutte queste modifiche richiedono una definizione più solida di OSINT, poiché i dati open source possono assumere molte forme.

In considerazione della natura mutevole delle informazioni pubblicamente disponibili, si considera[3] il periodo attuale come la seconda generazione di OSINT. L'ascesa del personal computer negli anni '90, periodo in cui è stato coniato l'acronimo OSINT, ha avuto un enorme impatto. Si può datare il passaggio all'OSINT di seconda generazione al 2005, anno in cui l'IC ha creato l'Open Source Center.[5] Anche Internet stava cambiando durante questo periodo, con la maggior parte dei contenuti online che si spostava su pagine

Web dinamiche, contenuti generati dagli utenti e social media. Questa transizione è spesso descritta come l'emergere del Web 2.0.[6]

Un punto di svolta nella concezione moderna dell'OSINT è stata la nascita del "Movimento Verde" in Iran nel 2009[3]. A seguito delle elezioni presidenziali in Iran del 2009, manifestanti oppositori di Mahmoud Ahmadinejad hanno chiesto le sue dimissioni, accusandolo di brogli elettorali.[23] Milioni di giovani iraniani si sono affidati a Internet per coordinare le loro attività, condividere contenuti virali e incoraggiare altri a partecipare alle manifestazioni. Per la prima volta, Internet è stata inondata di informazioni ai cittadini su un importante evento politico, in gran parte grazie alla combinazione di smartphone, connessioni Internet e social media. Durante la prima settimana delle proteste, il 60% di tutti i link ai blog pubblicati su Twitter riguardavano la politica iraniana.[24]

Segnali del carattere digitale di questa protesta sono il balzo della percentuale di utilizzo di Internet in Iran, dal 34% nel 2008 al 48% nel 2009[25], e degli abbonamenti di telefonia cellulare, dal 59% al 72% della popolazione.[26] A quel tempo, la BBC pubblicò un articolo intitolato "Internet brings events in Iran to life"[27], in cui affermava che nel paese stesse prosperando una nuova forma di "giornalismo cittadino". Durante le proteste il Washington Post ha condotto una sessione di Q&A con il sociologo e giornalista bielorusso Evgeny Morozov: [28]

> **Fairfax, Virginia**: Si è scritto molto sulla copertura in Iran lo scorso fine settimana e sul fatto che le testate giornalistiche statunitensi non hanno davvero avuto il loro peso, ma che Twitter e altri siti Web simili hanno diffuso notizie e fatto sapere alle persone in tutto il mondo cosa stava succedendo in Iran. Commenti?
>
> **Evgeny Morozov**: Abbiamo visto parecchi cittadini giornalisti fare un ottimo lavoro scattando foto e video delle proteste a Teheran quasi in tempo reale. Hanno, infatti, riempito una nicchia importante. Social network come Twitter, allo stesso modo, hanno svolto un ruolo importante nell'attirare l'attenzione delle persone su questi contenuti generati dagli utenti. Quindi, Flickr ha fornito ottime foto e Twitter ha prestato grande attenzione a queste foto. Ci sono state, infatti, molte critiche alla mancanza di copertura relativa all'Iran sulla CNN; Gli utenti di Twitter hanno persino organizzato un'intera campagna per affrontare questo problema chiamato #cnnfail. Penso che abbiano avuto successo: i dirigenti/reporter della CNN alla fine hanno dovuto rispondere a domande al riguardo.

Per la prima volta, qualsiasi individuo in tutto il mondo ha potuto estrarre da questi social network contenuti di livello di intelligence e scrivere articoli, previsioni e fornire analisi approfondite. Nonostante le proteste alla fine non abbiano avuto successo e il regime iraniano ha rapidamente ha notevolmente ristretto l'accesso a Internet e censurato qualsiasi forma di media d'accordo con l'opposizione, possiamo considerare la Rivoluzione Verde come un evento seminale nella storia dell'intelligence open source.

Con il senno di poi, non sorprende che l'Iran sia stato il luogo in cui l'OSINT moderno sia nato, difatti, prima della chiusura totale di Internet nel paese nel 2019, l'Iran aveva più utenti online di Bahrain, Emirati Arabi Uniti, Israele, Arabia Saudita, Siria, Yemen e Giordania messi insieme.[29] Durante la Rivoluzione, Matthew Weaver del Guardian espresse sorpresa per le realtà di questo nuovo mondo; "Quello che la gente dice a un certo punto della giornata viene poi confermato da fonti più convenzionali quattro o cinque ore dopo".[24]

Negli ultimi anni i giornalisti cittadini e i singoli investigatori (piuttosto che i governi) hanno guidato lo sviluppo dell'OSINT per via delle caratteristiche che ha come disciplina: è un campo che cambia a grande velocità, con nuovi strumenti e tecniche che vengono affinati e creati ogni giorno. Gli individui non vincolati da burocrazia possono rapidamente diventare utenti OSINT e creare report su informazioni prima impossibili.

Appena un anno dopo la Rivoluzione Verde, le rivoluzioni alimentate dai social media si sono diffuse in tutto il mondo arabo. La combinazione di rabbia pubblica, smartphone e social media ha scosso le dittature in

[3] https://www.treccani.it/enciclopedia/onda-verde/

tutto il Nord Africa e il Medio Oriente. Secondo l'allora direttore del OSC, l'organizzazione "non è stata in grado di prevedere lo sviluppo preciso dell'attivismo sociale basato su Internet nel mondo arabo".[30]

Tuttavia, negli ultimi anni gli Stati Uniti e altri hanno chiaramente preso atto. Nell'ottobre 2015, l'OSC è stata ribattezzata Open Source Enterprise (OSE) ed è stata trasferita nel Directorate for Digital Innovation (DDI) appena creato all'interno della CIA.[31] Inoltre, l'Office of the Director of National Intelligence (ODNI) riconosce che "le responsabilità di raccolta open source sono in generale distribuite in tutto l'IC", descrivendo l'OSC come uno dei principali, ma non l'unico, raccoglitore di OSIF.[1]

Nel settore privato, decine di società di intelligence open source sono nate sia nel Regno Unito che negli Stati Uniti, al servizio di una vasta gamma di clienti del settore pubblico e privato. Più di recente, come riportato da WIRED Magazine nell'aprile 2016, un'intera gamma di agenzie governative sta cercando modi per comprendere meglio la ricchezza di dati aperti disponibili online, rendendosi conto che OSINT può creare o interrompere operazioni.[32]

La maggior parte delle comunità di intelligence ha tardato ad apprezzare il valore di Internet per due motivi:

1. Le agenzie di intelligence cercano un vantaggio informativo attraverso la gestione dei segreti. Affidarsi a informazioni aperte e alle rispettive restrizioni di copyright è contrario a questa idea.

2. Nella maggior parte dei casi è più difficile, rischioso e costoso applicare metodi clandestini per acquisire fonti segrete, dando così l'impressione che tali fonti debbano avere un valore superiore rispetto alle fonti aperte, confondendo il metodo con il prodotto o scambiando la segretezza per conoscenza.

3 Definire OSINT

L'Office of the Director of National Intelligence degli Stati Uniti (ODNI) definisce sei discipline di raccolta o sei fonti di intelligence di base.[1] Tra esse è incluso anche l'OSINT, ma altre fonti non la considerano come una disciplina di intelligence (INT). Alcuni rifiutano OSINT perché le informazioni da fonti Open Source non vengono raccolte clandestinamente. Mark Lowenthal, ex Vicepresidente del National Intelligence Council, sostiene invece che l'OSINT non sia un'disciplina di intelligence in sé, ma una sfaccettatura delle altre varie discipline di intelligence.[33] L'intenzione di Lowenthal è quella di porre l'accento sul tipo di informazioni raccolte, piuttosto che sulle modalità della loro acquisizione per definire le discipline di intelligence. Stephen Mercado in "Studies in Intelligence" descrive in modo simile "corrispettivi aperti delle attività segrete quali [intelligence su] fotografie aeree (IMINT) e intelligence di segnali (SIGINT)."[34] Il modo in cui vengono definite le discipline di intelligence è importante, perché le definizioni spesso dettano in che modo le informazioni vengono trattate dagli analisti di intelligence all-source, in particolare come vengono valutate per credibilità e validità. Le definizioni influiscono anche nel determinare se un prodotto di intelligence verrà considerato come single-source o all-source, ciò ha impatto sulle valutazioni dell'IC riguardo l'affidabilità del prodotto. Inoltre, la definizione delle priorità degli sforzi di raccolta è spesso il modo in cui l'intelligence viene classificata (e quindi condivisa e diffusa) spesso deriva dalle definizioni delle discipline di intelligence, sottolineando quanto sia critico caratterizzare OSINT.

Questi problemi di definizione sono aggravati dal fatto che alcuni professionisti dell'intelligence tendono a pensare alle discipline di intelligence come uniche e distinte l'una dall'altra. Una visione d'insieme più efficace sarebbe pensare queste discipline come sovrapponibili, come fanno le definizioni di Lowenthal e Mercado. Le discipline di intelligence sono di rado completamente indipendenti l'una dall'altra e, nel contesto governativo, le definizioni di queste ultime sono a volte dettate più dalle autorità normative specifiche delle agenzie di raccolta dei dati di intelligence che da differenze distinte tra i metodi di raccolta o il materiale stesso. Ad esempio, la linea di demarcazione tra l'intelligence geospaziale (GEOINT) e la Measurement and Signature Intelligence (MASINT) è spesso confusa e lo sta diventando sempre di più grazie alle nuove tecniche di acquisizione delle immagini. Allo stesso modo, l'intelligence elettronica (ELINT) è talvolta considerata MASINT e talvolta una componente della SIGINT. La HUMINT è talvolta caratterizzata in modo simile alla SIGINT.

L'OSINT può fornire casi molto complessi e frequenti di questi tipi di sovrapposizione disciplinare ma non è l'unica tra le discipline di intelligence a farlo. Sempre più spesso, la GEOINT è anche OSINT, poiché i satelliti commerciali sono ora in grado di fornire una capacità di immagini dall'alto pari a quella storicamente fornita solo dalle piattaforme di raccolta riservate. L'OSINT derivata dai social media potrebbe essere considerata un tipo di HUMINT e SIGINT. Come nella HUMINT, la raccolta di dati sui social media fornisce approfondimenti e prospettive su un individuo che può fornire un punto di vista rappresentativo per una comunità o uno specifico per la popolazione nazionale. Analogamente alla SIGINT, la raccolta di dati sui social media può comportare la raccolta elettronica di un numero massiccio di registrazioni che vengono vagliate con mezzi tecnici per identificare le interazioni o le comunicazioni di interesse critico. Inoltre, poiché le informazioni su Internet sono sempre più protette - e la crittografia diventa sempre più pervasiva, accessibile e robusta - le informazioni che solo pochi anni fa erano apertamente disponibili al pubblico possono ora essere accessibili solo utilizzando metodi di raccolta clandestini o nascosti. L'implicazione di questa tendenza è che un numero maggiore di OSIF potrebbe essere più difficile da ottenere attraverso metodi di raccolta open-source. La Figura 1 fornisce una rappresentazione visiva di queste sovrapposizioni di discipline di intelligence. Non ha la pretesa di indicare tutti i modi in cui le discipline di intelligence potrebbero potenzialmente sovrapporsi; piuttosto, illustra le linee di demarcazione tra le diverse aree di intelligence.

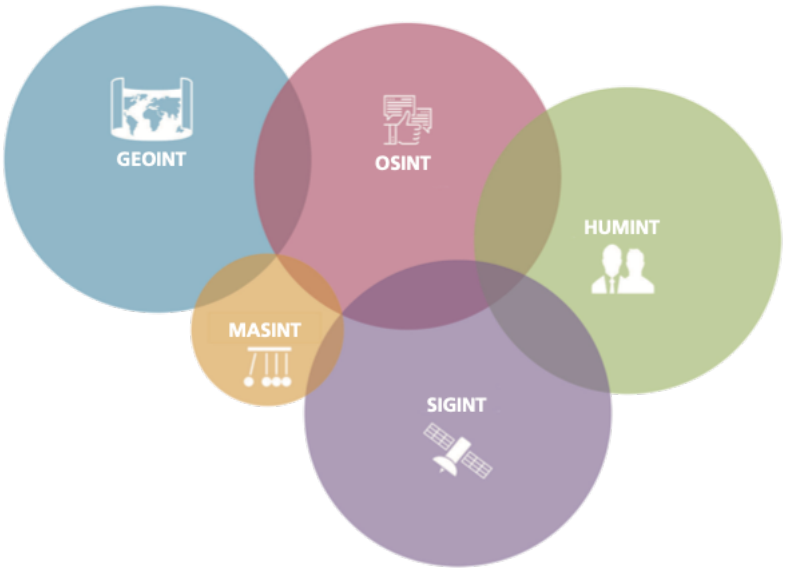


Figura 1: La Natura Sovrapponibile delle Discipline di Intelligence[3]

3.1 Definizione di Open Source e OSINT

Definiamo OSINT come il processo di raccolta d'informazioni attraverso la consultazione di fonti di pubblico dominio definite anche "fonti aperte". Ciò è coerente con la definizione statunitense contenuta nella sezione 931 della Public Law 109-163, che definisce l'OSINT come "intelligence prodotta da informazioni disponibili al pubblico e raccolta, sfruttata e diffusa tempestivamente a un pubblico appropriato allo scopo di rispondere a una specifica esigenza di intelligence".[35] L'OSIF è semplicemente un dato non riservato disponibile al pubblico, mentre l'OSINT è il risultato dell'elaborazione e dello sfruttamento delle informazioni per convalidarne la rilevanza, l'accuratezza e l'utilizzabilità da parte dei consumatori.

Ciò che costituisce OSINT e OSIF non è universalmente definito. Ad esempio, l'Handbook of Intelligence Studies, un popolare libro di testo sull'intelligence, identifica quattro categorie distinte di OSIF e OSINT. I dati open-source sono definiti come informazioni grezze provenienti da fonti primarie, quali stampa, radiodiffusione, colloqui orali o altre forme di informazioni. Le OSIF sono descritte come dati che possono essere messi insieme a partire da informazioni generiche che sono in genere ampiamente diffuse; le fonti includono giornali, libri, trasmissioni e report quotidiani. OSINT è definita come un'informazione che è stata "deliberata, scoperta, discriminata, distillata e diffusa a un pubblico selezionato".[36] L'"OSINT convalidata" si distingue dall'OSINT per l'elevato grado di validità e certezza ad essa associato. Heather J. Williams e Ilana Blum in "Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise" propongono la seguente classificazione:

- I dati open-source sono materiale che, se isolato, avrebbe scarso valore individuale, ma che ha un valore di intelligence se elaborato. Ad esempio, un singolo tweet che riflette l'opinione di un individuo casuale sullo Stato Islamico dell'Iraq e al-Sham (ISIS) non ha quasi alcun valore di intelligence; tuttavia, sintetizzare tutti i tweet sulle opinioni sull'ISIS all'interno di un'area geografica ha un grande valore di intelligence. Allo stesso modo, i singoli indirizzi di protocollo (IP) non hanno alcun valore di intelligence, ma la mappatura dei 4,3 miliardi di indirizzi IP del mondo fornisce un quadro globale dell'utilizzo di Internet. [37] I dati open source includono materiale pubblico che non è esplicitamente pubblicato, ma è comunque disponibile pubblicamente o commercialmente, come le immagini satellitari commerciali.

- L'OSIF è materiale che può essere ottenuto legalmente tramite richiesta, acquisto o osservazione da parte di un membro del pubblico. [38] L'OSIF è quindi la categoria più ampia di informazioni disponibili pubblicamente o commercialmente.

[3]

3.2 Sottotipi OSINT

Una sfida nel definire l'OSIF è rappresentata dal fatto che esistono poche sottocategorie riconosciute per distinguere i vari tipi di informazione e le definizioni esistenti non colgono con precisione la natura mutevole delle informazioni pubbliche. La definizione di "letteratura grigia", ad esempio, è stata particolarmente complicata dall'avvento di Internet. Nel 1995, l'Interagency Gray Literature Work Group del governo degli Stati Uniti ha definito la letteratura grigia come "materiale open-source straniero o nazionale che di solito è disponibile attraverso canali specializzati e non può entrare nei normali canali o sistemi di pubblicazione, distribuzione, controllo bibliografico o acquisizione da parte di librerie o agenti di sottoscrizione". [39] Questa definizione potrebbe includere un'ampia gamma di informazioni, come la letteratura grigia e la letteratura di massa. "Questa definizione potrebbe includere un'ampia gamma di tipi di informazioni: documenti di conferenze, documenti aziendali, tesi di laurea, rapporti governativi, newsletter, letteratura commerciale, relazioni di viaggio, tipicamente pubblicati da istituti di ricerca, governi nazionali, editori privati, aziende, associazioni e sindacati di categoria, think tank e università. L'aspetto più difficile del trattare la letteratura grigia in passato era la sua accessibilità, ed essa continua a essere alla base delle definizioni attuali.[40] Tuttavia, gran parte di questo materiale è ora disponibile online. Concentrarsi sull'accesso, quindi, non è più un criterio efficace per definire la letteratura grigia.

Nel 2007 il Congressional Research Service ha descritto quattro categorie di OSIF: "dati e informazioni ampiamente disponibili; dati commerciali mirati; singoli esperti; letteratura 'grigia'".[41] Queste categorie, tuttavia, non sono coerenti con la concettualizzazione dell'IC dell'open source, né catturano efficacemente i contenuti dei social media. C'è ovviamente un pericolo nel definire l'OSIF in modo troppo restrittivo, dato il rapido cambiamento della natura delle fonti e delle piattaforme online. Tuttavia, senza un quadro di riferimento per differenziare le ampie fasce di OSIF, il ciclo di intelligence OSINT non può essere definito con precisione a causa delle notevoli differenze nell'elaborazione e nello sfruttamento dei diversi tipi di OSIF.

Williams e Blum propongono di dividere le OSIF in quattro categorie: due categorie principali, ciascuna biforcata a un ulteriore livello. La prima differenziazione è determinata dal generatore dei contenuti: contenuti generati dalle istituzioni e contenuti generati dagli individui. I contenuti generati dalle istituzioni sono costituiti dai media e da altri contenuti istituzionali, molti dei quali possono essere stati precedentemente definiti come letteratura grigia. I contenuti generati dagli individui, o i contenuti dei social media, si dividono in long-form e short-form, che presentano differenze importanti per l'elaborazione e l'utilizzo.

3.2.1 Contenuto dei media di informazione

Il contenuto dei mezzi di informazione è auto-identificato e riconosciuto pubblicamente come giornalismo. Le sue fonti sono giornali multimediali, riviste (sia cartacee che online), televisioni e radio. I media includono anche siti aggregatori di notizie, che possono pubblicare o meno contenuti originali. I contenuti dei mezzi di informazione includono i contenuti prodotti dallo Stato quando sono distribuiti specificamente da un organo di informazione.

3.2.2 Letteratura grigia

La letteratura grigia è costituita da contenuti provenienti da istituzioni e organizzazioni non mediatiche, sia pubbliche che private. Include materiale proveniente da istituti di ricerca, governi nazionali, editori privati,

aziende, associazioni di categoria e sindacati, think tank e università. L'ipotesi di fondo è che la maggior parte dei contenuti istituzionali non esista solo nello spazio virtuale, ma che in genere vi sia una certa presenza e coesione istituzionale. Nonostante gli sforzi iniziati decenni fa per organizzare meglio l'acquisizione, l'archiviazione a lungo termine e la distribuzione della letteratura grigia, questa viene ancora spesso raccolta e utilizzata ad hoc.

3.2.3 Contenuti long-form sui social media

I contenuti long-form di singoli utenti sono materiale ricco di testo proveniente da singoli individui o da piccoli gruppi. Include materiale proveniente da blog e siti come Reddit e Tumblr. Gran parte dell'analisi dei contenuti dei social media si è concentrata sui contenuti in short-form, lasciando i contenuti lunghi spesso poco utilizzati.

3.2.4 Contenuti short-form sui social media

I contenuti brevi di singoli utenti sono materiale proveniente da piattaforme come Facebook, Twitter e LinkedIn. A differenza dei contenuti di long-form, quelli in short-form hanno generalmente uno scarso valore di intelligence individualmente; il valore di intelligence si ottiene dall'aggregazione di tali informazioni. Eccezion fatta quando i contenuti short-form sui social media sono ottenuti da account specifici di grande interesse, ad esempio gli account di persone famose come alti esponenti del governo, leader di pensiero e giornalisti di spicco. Gli short-form di alto valore potrebbe essere raccolto da account di individui che fanno parte di un gruppo preso di mira dalla IC, come un'unità militare speciale o un gruppo militante.

3.3 Metodologie OSINT: Il Ciclo delle Operazioni OSINT

Solo una parte dell'enorme volume di OSIF che viene diffuso e condiviso quotidianamente si qualifica come informazione rilevante, tempestiva e fruibile per un analista OSINT.[42] Determinare cosa sia meno o più rilevante richiede un'enorme quantità di lavoro distribuito sull'intero spettro dell'intelligence, dalla raccolta iniziale alla diffusione dei risultati al decisore politico che li riceve. La trasformazione delle informazioni dall'intelligence grezza comporta fasi cruciali per fornire il contesto in cui valutare la validità e l'affidabilità di un rapporto.

L'OSINT, tuttavia, ha ancora bisogno di una metodologia chiara.[33] Esistono diversi modelli per descrivere la metodologia dell'intelligence. Al fine di trasformare i dati grezzi in intelligence azionabile, la Intelligence Community ha derivato un modello chiamato ciclo dell'intelligence[43]. Esso descrive questo processo come pianificazione e direzione, raccolta, elaborazione, analisi e produzione e diffusione. Viene applicato a tutte le fonti di intelligence e, in particolare, all'OSINT. Questo modello è stato adottato da Gibson[44] e, con alcuni aggiustamenti, anche da Hassan[45]. Bazzell presenta un'interpretazione pratica che viene utilizzata come manuale di formazione obbligatorio dalle agenzie governative statunitensi[46]. Altri lavori enfatizzano la raccolta e l'analisi delle informazioni e, pertanto, introducono modelli incentrati su questi compiti. Ciò vale per il modello completo in tre fasi derivato da Pastor, et al.[47], nonché per il modello di Tabatabaei, et al.[48].

Nel contesto dell'OSINT, ci si può concentrare su quattro fasi chiave: raccolta, elaborazione, sfruttamento e produzione, come mostrato nella Figura 2 (l'elaborazione e lo sfruttamento possono non avvenire in modo completamente sequenziale, ma piuttosto in parallelo o di pari passo). In termini più semplici, queste fasi possono essere descritte come l'acquisizione di informazioni, la convalida di tali informazioni, l'identificazione del valore delle informazioni e la fornitura delle informazioni ai clienti. Nelle sezioni che seguono, si scomporranno ognuna di queste aree in componenti in modo specifico per ogni tipo di OSINT piuttosto che generalizzare la fase per tutti i tipi di materiale open-source.

La difficoltà di ogni componente del ciclo metodologico è stata classificata da Williams e Blum generalmente come facile, media o difficile per ciascuno dei sottotipi di OSINT (cfr. Figura 3).[3] Nel caratterizzare la

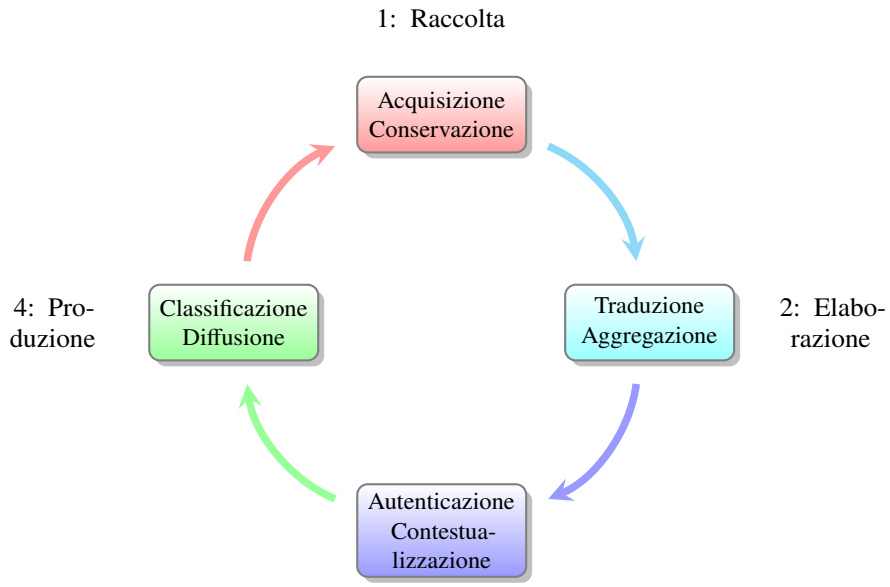


Figura 2: Il Ciclo delle Operazioni Osint[3]

difficoltà, sono stati considerati i fattori che contribuiscono al livello di difficoltà, come le ore di lavoro, le risorse informatiche necessarie e l'accesso a obiettivi difficili. Questo tipo di classificazione non ha valore universale. Ad esempio, alcuni analisti potrebbero sostenere che l'acquisizione dei contenuti dei social media è abbastanza facile, data la facilità di accesso ai dati di Twitter. Tuttavia, molti degli obiettivi difficili della IC utilizzano piattaforme nazionali e l'uso di Twitter è vietato o limitato, per cui estrarre i contenuti dei social media solo da questa piattaforma non risponderebbe a molte esigenze della IC.

3.3.1 Raccolta

La prima fase, la raccolta, comprende l'identificazione delle informazioni potenzialmente utili e la conservazione di tale materiale. Questa fase richiede una guida, esplicita o generale, che consenta ai raccoglitori di fonti aperte di identificare i tipi di informazioni da raccogliere e di dare priorità agli sforzi di raccolta per riflettere i requisiti della ricerca. L'acquisizione è la raccolta fisica o elettronica di queste informazioni. La conservazione (Retention) è il mantenimento delle OSIF acquisite.

Dei quattro tipi di OSIF qui considerati, i contenuti dei media sono i più facili da raccogliere. Per l'OSINT di prima generazione, l'acquisizione fisica dei dati trasmessi dai media presentava sfide logistiche che richiedevano all'FBIS di spostarsi in diverse località geografiche per intercettare le trasmissioni. La raccolta di materiale stampato dipendeva dalla presenza di un ufficiale diplomatico o di un raccoglitore clandestino per acquisire fisicamente il materiale pubblicato. Oggi, tuttavia, con la maggior parte delle informazioni dei media disponibili online, le sfide logistiche si sono spostate dall'elaborazione alla gestione delle informazioni. La conservazione delle informazioni dei media è abbastanza semplice. Il volume di tali informazioni è gestibile e le informazioni vengono generalmente fornite in un formato standardizzato e di testo.

La letteratura grigia, come i contenuti dei media, sta diventando più facile da raccogliere, per ragioni simili. I creatori di letteratura grigia sono stati più lenti dei media nel passare ai contenuti online, quindi ci sono ancora casi in cui un raccoglitore deve acquisire fisicamente le informazioni in formato cartaceo, in particolare nei paesi in via di sviluppo, dove l'uso di Internet da parte delle istituzioni potrebbe non essere diffuso. Come nel caso dei contenuti dei media, la conservazione della letteratura grigia non è molto difficile.

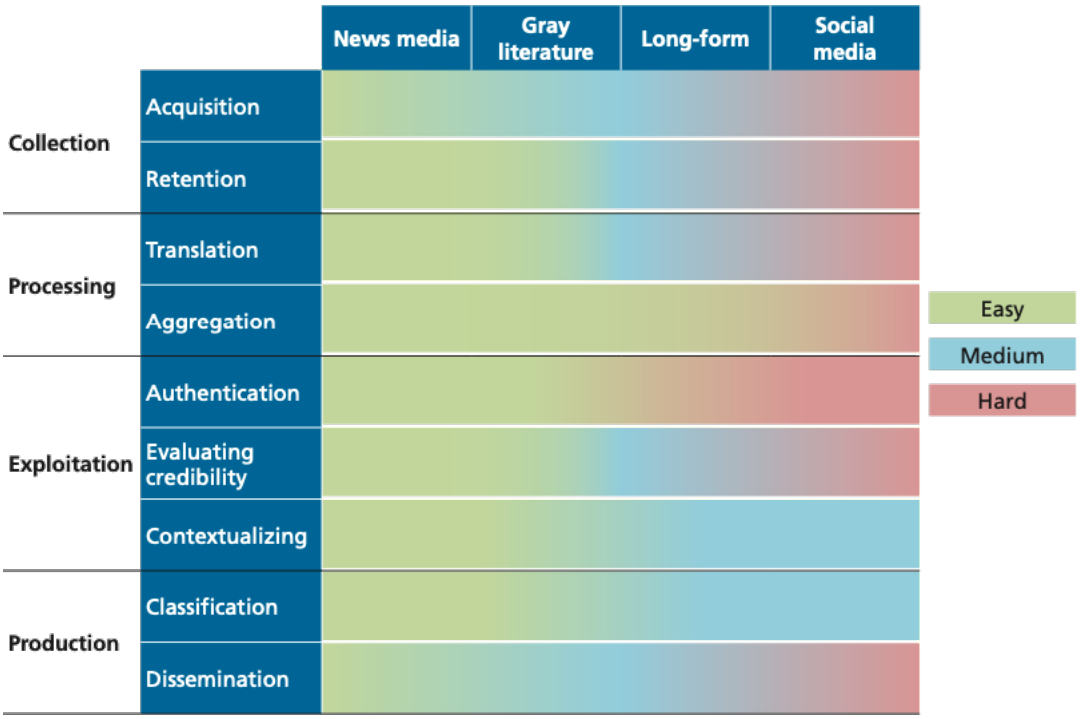


Figura 3: Difficoltà dei passaggi del ciclo OSINT, per tipo di OSIF[3]

Le informazioni sui social media, al contrario, presentano molte sfide uniche nella fase di raccolta, sia per i contenuti in short-form che in long-form. Innanzitutto, un quadro completo dei dati grezzi può essere difficile da acquisire. Nella fase di avvio dell'analisi dei contenuti dei social media, l'analisi dei social media era facilmente accessibile e talvolta persino gratuita per l'uso. Tuttavia, l'analisi dei social media è diventato un settore consolidato, quindi piattaforme come Topsy4.1 sono state acquistate e chiuse da aziende più grandi che cercano di monetizzare in questi mercati. L'aggregazione sui social delle aziende che commercializzano i dati dei social media spesso fornisce solo una frazione dei dati da una piattaforma di social media o un set di dati solo da una specifica finestra temporale.

Inoltre, questi fornitori tendono anche a concentrarsi sui dati dei social media dalle piattaforme con sede negli Stati Uniti, principalmente Twitter e Facebook. Oltre a questo, anche se l'IC può acquisire un set completo di dati multimediali dai social piuttosto che un sottoinsieme, i dati non presentano un campione rappresentativo per una popolazione. I gruppi demografici non utilizzano i social media in modo uniforme e in molte località di interesse per l'IC, l'utilizzo può essere enormemente influenzato dalla classe socioeconomica.

Inoltre, i contenuti dei social media, sia quelli di lunga durata che quelli di breve durata, sono più dinamici di quelli dei media d'informazione o della letteratura grigia. Un articolo di cronaca (ad eccezione di eventuali correzioni) non è generalmente un documento vivente: se una storia è cambiata, verrà generato un nuovo articolo separato. Al contrario, una argomento di discussione di tendenza può suscitare interesse e aggiornamenti per pochi giorni o settimane, oppure può continuare per anni. L'acquisizione e la conservazione dei contenuti dei social media, in particolare, devono essere costanti e in tempo reale, poiché i contenuti d'impatto possono essere pubblicati e rimossi in un breve periodo di tempo se suscitano polemiche o rivelano informazioni sensibili, casi che potrebbero essere di particolare interesse per la IC. Infine, sia i contenuti long-form che quelli short-form dei social media sono sempre più spesso presentati in formati diversi da quello

testuale. I video di YouTube sono un esempio di contenuti social media di long-form in un formato diverso, mentre i dati social media di short-form in un formato non testuale includono immagini su piattaforme come Flickr e video "live" su piattaforme come Facebook e Twitter.

Bazzell fornisce il resoconto più dettagliato di questa fase strutturandola in tre fasi[46]. In primo luogo, suggerisce di ricorrere a motori di ricerca, siti Web e servizi specializzati che potrebbero richiedere un pagamento o una commissione. In secondo luogo, l'indagine prosegue con una ricerca web iniziale degli identificatori seguita dall'utilizzo di applicazioni, strumenti e tecniche OSINT selezionati a seconda dell'obiettivo dell'indagine. Per strutturare questo utilizzo, deriva un flusso di lavoro per ciascuno degli identificatori di indirizzo e-mail, nome utente, nome reale, numero di telefono, nome di dominio e posizione. Ogni flusso di lavoro inizia con un determinato identificatore e propone percorsi diversi inclusi strumenti specifici. Ciò si traduce in nuove informazioni che possono essere ulteriormente sfruttate per ottenere maggiori informazioni. Ad esempio, un determinato nome utente identificatore è potenzialmente utile per identificare il nome reale, l'indirizzo e-mail o il profilo di un social network. Il flusso di lavoro include diversi approcci su come procedere. Un approccio è un controllo manuale per tutti i social network per il nome utente dato, identificando così potenzialmente il vero nome. Un altro percorso descritto nel flusso di lavoro è l'ipotesi dell'indirizzo e-mail in base alle informazioni fornite. Un terzo percorso prende il nome utente come input in una serie di strumenti forniti da Bazzell. Inoltre, l'input viene elaborato da motori di ricerca web standard e specializzati e ulteriormente arricchito con informazioni provenienti da database compromessi. Tutti questi flussi di lavoro, tuttavia, sono nella maggior parte dei casi adattati a una ricerca situata negli Stati Uniti. Nella terza e ultima fase della fase di raccolta, vengono acquisiti tutti i reperti.

3.3.2 Elaborazione

La seconda fase, l'elaborazione, consiste nel convalidare le informazioni e renderle fruibili. L'elaborazione può assumere diverse forme, tra cui la traduzione dei materiali e la trasformazione di materiali video o fotografie in intelligence utilizzabile. L'elaborazione nell'OSINT di seconda generazione presenta un cambiamento radicale alla generazione precedente, sia per le modifiche ai metodi esistenti che per i requisiti dei nuovi metodi. Molti dei compiti svolti nell'elaborazione possono ora essere svolti più facilmente e a costi inferiori grazie all'uso di programmi software, tra cui le versioni professionali di Google Translate. Allo stesso tempo, l'OSINT dispone di un'abbondanza di informazioni disponibili in un formato meno strutturato, il che rende l'elaborazione molto più complessa. Identifichiamo due componenti dell'elaborazione: la traduzione e l'aggregazione. Queste componenti non devono necessariamente avvenire in una determinata sequenza, anche se in alcuni casi una potrebbe aiutare l'altra.

Traduzione L'elaborazione dei dati dei media di informazione comporta principalmente la traduzione. Se un tempo quest'ultima costituiva la parte più consistente dello sforzo dell'FBIS, è stata radicalmente influenzata dal rapido progresso della traduzione automatica. Sebbene i linguisti abbiano ancora un ruolo critico da svolgere - fornendo sfumature e contesto culturale al materiale in lingua straniera - possono ora concentrare i loro sforzi sulla fornitura di valore analitico nella fase di sfruttamento. La traduzione automatica è più efficiente per i contenuti dei media, che utilizzano un vocabolario standard e spesso seguono una struttura formulaica. Anche la letteratura grigia segue generalmente standard di scrittura professionale che sarebbero favorevoli alla traduzione automatica, sebbene gli argomenti avanzati e specifici trattati nella letteratura grigia richiedano talvolta l'intervento umano. Le informazioni dei social media presentano vantaggi e svantaggi per la traduzione automatica. Da un lato, i post sui social media tendono a contenere un numero limitato di caratteri: Twitter, ad esempio, limita gli utenti a 280 caratteri. D'altro canto, è più probabile che i post sui social media contengano espressioni gergali, abbreviazioni o emoji. Inoltre, possono utilizzare più lingue e probabilmente contengono più spesso errori tipografici. Mentre i contenuti dei social media in long-form possono contenere informazioni sufficienti a fornire una traccia coerente attraverso la quale dedurre lo stile o la

posizione dell'autore, i contenuti dei social media in forma breve hanno meno probabilità di fornire tale traccia, a meno che non venga compilato un corpus del materiale o dell'attività.

Aggregazione L'aggregazione, che in genere non è necessaria per i dati dei mezzi di informazione e della letteratura grigia, è una fase critica per l'analisi di molti tipi di contenuti dei social media, in particolare di quelli di breve durata. L'aggregazione può anche comportare una riduzione o un'integrazione nel tradurre un insieme di dati in una forma utilizzabile. Molte aziende commerciali forniscono servizi di aggregazione dei dati che eliminano la necessità di una raccolta diretta da parte dell'IC. Se da un lato questi aggregatori di dati possono ridurre al minimo la raccolta e l'elaborazione delle informazioni, dall'altro potrebbero non fornire dati da più piattaforme e non fornire campioni completi di dati. Può anche essere difficile sapere esattamente quali dati sono stati inclusi nel set di dati, il che complica la sua capacità di autenticare i dati e inserirli in un contesto appropriato.

3.3.3 Sfruttamento

Lo sfruttamento (talvolta definito anche analisi) cerca di determinare il valore delle informazioni. Come osserva l'ex ufficiale della CIA e studioso di intelligence Arthur Hulnik, una delle sfide più significative associate all'uso dei prodotti OSINT è l'enorme volume di informazioni disponibili pubblicamente e i gradi di affidabilità insiti in tali informazioni. Pertanto, nell'analisi dell'OSINT occorre dedicare molto tempo a separare l'intelligence affidabile e "buona" da quella "cattiva".[49] Gli operatori devono essere in grado di "raccogliere, giudicare e smistare le informazioni, conoscere e gestire le restrizioni e comprendere i diversi utenti, le esigenze, i compiti, il complesso delle informazioni, l'organizzazione, le istituzioni e la legge".[42] Il prodotto finito dovrebbe fornire conclusioni analitiche guidate dalle fonti disponibili.

Lo sfruttamento si suddivide quindi in tre fasi: autenticazione, valutazione della credibilità e contestualizzazione.

Autenticazione L'autenticazione cerca di verificare se le informazioni sono ciò che dicono di essere. Per le informazioni provenienti da fonti istituzionali, questo è abbastanza semplice. Gli articoli sul New York Times hanno un'alta probabilità di essere stati pubblicati consapevolmente e intenzionalmente dal New York Times. Allo stesso modo, la letteratura grigia pubblicata dai siti web governativi può essere ritenuta con elevata certezza prodotta e diffusa dal governo. L'autenticazione dei contenuti dei social media è molto più difficile. Gli utenti possono mascherare di proposito la loro vera identità o fornirne una falsa. Questo va oltre il semplice nome dell'utente. Ad esempio, una persona potrebbe essere disonesta riguardo alla sua posizione o alle sue caratteristiche personali. Se si cerca di accertare l'atmosfera all'interno di un Paese, è molto importante che gli utenti si trovino all'interno di quel Paese e non siano membri di una diaspora. L'autenticazione può essere necessaria in concomitanza con le operazioni di aggregazione dei dati per garantire che un campione o un composito di dati non sia distorto.

Valutazione La valutazione della credibilità, come l'autenticazione, è abbastanza semplice per i contenuti dei media tradizionali e della letteratura grigia, ma estremamente difficile per i contenuti dei social media. Una misura di credibilità cerca di determinare se le informazioni sono affidabili, cioè se sono state fornite senza l'intento di negare o fuorviare e se la loro fonte ha un accesso plausibile. Il New York Times, ad esempio, pubblica quasi sempre materiale con uno scopo preciso: desidera che i suoi contenuti siano accurati ed è trasparente sulle sue fonti. Questo può essere meno vero per le fonti mediatiche straniere, in particolare per i media gestiti dallo Stato, che hanno l'intenzione influenzare o inviare messaggi alle loro popolazioni. Tuttavia, probabilmente esiste una serie di materiale proveniente da tali fonti che potrebbe fornire qualche indicazione sulla credibilità delle loro informazioni.

A differenza dei mezzi di informazione, i social media non contengono generalmente informazioni di seconda mano. Il contenuto di solito proviene direttamente dalla fonte. Tuttavia, non è sempre così e l'originalità della fonte può essere sospetta. Retweet, repost e bot sono esempi di dati dei social media che si sono rivelati importanti per offuscare le intenzioni della fonte originale. Anche se una fonte fornisce informazioni su un evento di cui è stata testimone, possiamo fidarci del suo resoconto, dato che potremmo sapere poco dei pregiudizi e delle competenze della fonte? Con questo non si vuole dare per scontato che le persone interpretino sempre intenzionalmente i fatti in modo errato. Un esempio significativo è quello delle videocamere fatte indossare ai poliziotti, che secondo alcuni fanno luce sugli interventi della polizia, ma che possono essere di difficile interpretazione e che può essere soggetta a pregiudizi cognitivi.[50]

Contestualizzazione La contestualizzazione consente all'analista di fonti aperte di trasmettere la propria competenza in materia al consumatore finale. Ciò può comprendere commenti sulla fonte che forniscono informazioni aggiuntive, come quelle relative alla credibilità o anche comportare la compilazione di più elementi di OSIF da qualsiasi documento in un prodotto che fornisce un quadro più completo di una questione.

3.3.4 Produzione

Nella fase finale, la produzione, le informazioni vengono fornite a un consumatore in una forma utilizzabile. Il più delle volte si tratta di un analista di intelligence all-source che incorporerà il prodotto ricevuto in una produzione multi-intelligence. Tuttavia, un prodotto open-source può anche essere ad alta priorità o abbastanza completo da essere fornito direttamente a un legislatore o a un altro cliente dell'intelligence. Ciò è simile ad altre discipline di intelligence, in cui l'intelligence umana, dei segnali o geospaziale è generalmente incorporata in un prodotto analitico all-source, ma a volte viene fornita nella sua forma grezza direttamente a un cliente dell'intelligence.

La fase di produzione comprende anche l'assegnazione di un livello di segretezza a un prodotto OSINT. Anche se il prodotto può derivare da OSIF, i dettagli della raccolta, dell'elaborazione e dello sfruttamento di tali informazioni possono giustificare un livello di segretezza più elevato. Le OSIF possono soddisfare dei requisiti di informazioni segrete di intelligence, in particolare se combinate con altre informazioni. Ad esempio, le informazioni possono essere acquisite attraverso mezzi ufficiali o sensibili, quando l'esposizione del possesso delle informazioni ne comprometterebbe la continua accessibilità. I produttori di fonti aperte possono anche utilizzare tecnologie di elaborazione e sfruttamento classificate che giustificherebbero la classificazione delle informazioni.[51]

Diffusione Anche la diffusione è una componente della fase di produzione. L'analisi open source viene spesso diffusa sotto forma di rapporto scritto. Tuttavia, i prodotti possono anche assumere la forma di briefing verbali o visualizzazioni grafiche.

Il mezzo utilizzato per la diffusione è spesso il meccanismo di distribuzione più semplice, piuttosto che il più efficace. Video, audio o grafici interattivi possono spesso essere più efficaci dei rapporti scritti per trasmettere particolari informazioni. Gli analisti di intelligence all-source generalmente ottengono i loro rapporti di intelligence da un database testuale, come Trident, WISE o Pathfinder. Allo stesso modo, i consumatori di intelligence spesso ricevono i prodotti di intelligence in un libro di briefing stampato. Tuttavia, le maggiori capacità dei portali open-source e il passaggio del Presidential Daily Brief al formato iPad stanno aprendo la strada a meccanismi più creativi per la trasmissione delle informazioni, come visualizzazioni di dati e file dinamici.

Feedback Questa fase conclude l'indagine. Mentre la US National Intelligence e Gibson includono la valutazione del feedback per migliorare i propri processi[43][44], Bazzell conclude un'indagine con l'archiviazione dei risultati e un processo di pulizia[46]. Il resto dei modelli rinuncia a questa fase[47][48][45].

3.4 Dati, informazioni e intelligence

L'ultima sezione ha mostrato il processo di un'indagine OSINT non solo come acquisizione di dati, ma anche come trasformazione dei dati raccolti in informazioni e, infine, in intelligence. Questa sezione chiarisce i termini dati, informazioni e intelligence.

La differenziazione tra dati, informazioni e intelligence deriva dal NATO Open Source Intelligence Handbook ed è spesso inclusa nella discussione sulla metodologia delle indagini OSINT, ad esempio da Gibson o Hassan.[44][45] Questo processo di trasformazione dai dati all'intelligence è visualizzato nella Figura 4.

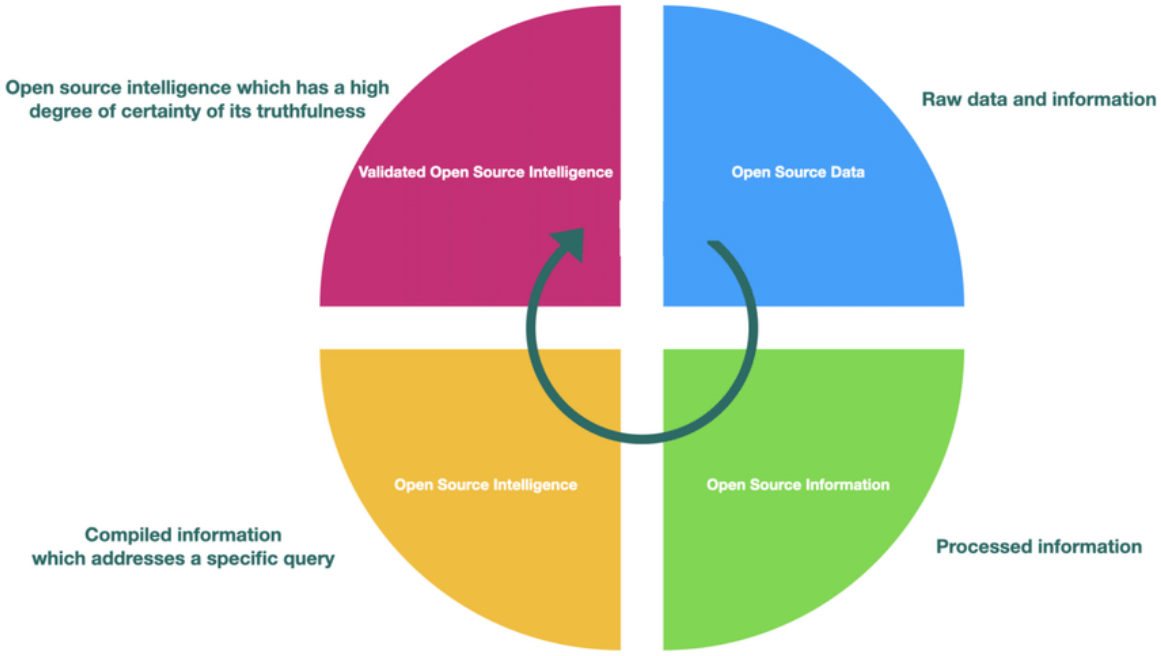


Figura 4: Visualizzazione dell'elaborazione dei dati durante un'indagine OSINT come descritto in Gibson

I dati descrivono l'output ottenuto durante la fase di raccolta nella Figura 2. Sono considerati come un insieme di fatti senza alcuna spiegazione o analisi.[45] I dati acquisiti possono essere classificati in base al loro formato. Si possono distinguere tra dati strutturati, dati semi-strutturati e dati non strutturati. I dati strutturati sono organizzati da un modello di dati sottostante e risultano in un formato standard facile da elaborare automaticamente. Un esempio importante di dati strutturati sono i database SQL. Sebbene non siano organizzati in modo così rigido come i dati strutturati, anche i dati semi-strutturati contengono alcuni elementi strutturali. Ne sono un esempio i documenti JavaScript Object Notation (JSON), Extensible Markup Language (XML) e Hypertext Markup Language (HTML). Siti web, relazioni, immagini, audio e video sono considerati dati non strutturati.[44] La Figura 4 rappresenta questo output come dati open source.

A seconda del modello sottostante, le informazioni sono il risultato della fase di raccolta o della fase di elaborazione della Figura 2. Sono prodotte dall'elaborazione dei dati raccolti. Si producono elaborando i dati raccolti. A seconda della natura dei dati, l'elaborazione comprende la traduzione, la decifrazione o la conversione del formato, oltre al filtraggio, alla correlazione, alla classificazione, al raggruppamento e all'interpretazione dei dati. La Figura 4 si riferisce a questo output come informazione open source.

La compilazione delle informazioni per rispondere a una domanda specifica dà origine all'intelligence. È il risultato dell'integrazione, della valutazione e dell'analisi delle informazioni durante la fase di analisi della Figura 2. Fino a questo punto, le fasi del ciclo dell'intelligence di cui alla Figura 2 corrispondono ai

diversi output di cui alla Figura 4. Tuttavia, la Figura 4 illustra un quarto e ulteriore tipo di output che non rientra nel ciclo dell'intelligence. Secondo la NATO, si tratta dell'intelligence di fonte aperta convalidata. È descritta come un'intelligence di fonte aperta a cui "può essere attribuito un alto grado di certezza"[44]. Ciò richiede la verifica e la convalida dell'intelligence open source derivata. Questo può essere fatto potenzialmente utilizzando altre fonti di intelligence.

La trasformazione dei dati può portare a un'abbondanza di tipi diversi di informazioni. Pertanto, si ripropone la classificazione proposta da Böhm e Lolagar per strutturare il tipo di informazioni che ci si può aspettare da un'indagine OSINT.[52]

Le informazioni sono classificate in base all'obiettivo di un'indagine OSINT. Sono stati identificati sette tipi di entità destinatarie. Le informazioni acquisite possono riguardare:

Singola Persona Le informazioni reperite vanno dalle informazioni sulla vita reale, come nome e cognome, indirizzo, occupazione o informazioni finanziarie, alla persona online con nome utente, indirizzo e-mail o presenza sui social media.

Gruppo di persone In particolare, le indagini penali spesso non si concentrano solo su una singola persona, ma su un insieme di persone per capire le loro relazioni e interazioni personali.

Organizzazione In questo documento, per organizzazione si intende un'entità con uno scopo definito e articolato che la distingue da una persona o da un gruppo di persone. Ne sono un esempio le aziende, le istituzioni, le associazioni o addirittura i Paesi. Il numero di informazioni interessanti comprende dettagli sulle operazioni commerciali, sulla pianificazione strategica, sulle relazioni con i clienti, sui dipendenti e infine sui clienti.

Sistema informatico Riassume tutte le informazioni relative ai sistemi informatici. Contiene informazioni sui nomi di dominio e sui sottodomini esistenti, sui software utilizzati e sulle rispettive versioni, nonché sulle porte aperte.

Evento L'osservazione delle interazioni tra persone può portare a informazioni su un evento che si svolge online o nella vita reale, con dettagli su data, luogo e partecipanti.

Posizione Include i dettagli relativi a un indirizzo fisico o a un insieme di coordinate.

Oggetto Copre tutti gli obiettivi non compresi in una delle classi precedenti. Include immagini e video e il loro contenuto.

Questa classificazione non è distinta, il che significa che un obiettivo può rientrare in più di una classe. Ad esempio, l'informazione "*Woodstock*" può essere classificata come gruppo di persone, evento o luogo. La sua classificazione dipende dal contesto dell'interrogazione iniziale, che influenza anche il modo in cui l'informazione viene ulteriormente elaborata.

3.5 Casi d'uso

L'OSINT non è più utilizzata esclusivamente dalla comunità dell'intelligence, ma da una serie di altre figure. Si vuole qui illustrare una serie di casi d'uso osservati negli ultimi anni in cui persone con background e motivazioni diverse hanno utilizzato diversi tipi di informazioni da fonti open source.

3.5.1 Intelligence

Nel 2015, un jihadista ha postato un selfie davanti a una fabbrica di bombe dello Stato Islamico, rivelando la struttura dell'edificio. 23 ore dopo, l'esercito statunitense ha lanciato un attacco distruggendo l'edificio.[53]

3.5.2 Giornalismo

Bellingcat è un collettivo di ricercatori, investigatori e "*citizen journalists*" che utilizza fonti open source e indagini sui social media per sondare una varietà di argomenti diversi con risultati impressionanti. Tra questi, l'identificazione di agenti dei servizi segreti russi come principali sospettati nelle indagini sul Volo Malaysia Airlines 17[54] e sull'avvelenamento della famiglia Skripal.[55] Inoltre, hanno fornito analisi sull'attacco chimico a Douma, in Siria,[56] e sull'uso di droni da parte di attori non statali in Siria e in Iraq.[57] Hanno smascherato un falso profilo sui social network attribuito a una persona inesistente ampiamente citato dai media ucraini e russi anti-Putin come funzionario del Pentagono[58] e hanno rivelato la spedizione illegale di precursori dell'agente nervino sarin in Siria da parte di aziende belghe.[59]

3.5.3 Forze dell'Ordine

A partire dal 2016, la German Police University di Münster ha condotto uno studio di ricerca per valutare come l'OSINT possa fornire informazioni rilevanti per le attività quotidiane di applicazione della legge e, quindi, ridurre i rischi per le forze di polizia e per il pubblico in generale.[60] Epple e Ludewig concludono che "implementando l'Open Source Intelligence nei centri di smistamento della polizia, si possono ottenere informazioni rilevanti per la missione e che l'OSINT è uno strumento adatto a garantire un compimento più professionale della missione, una migliore protezione della popolazione e una migliore protezione personale degli agenti di polizia".[60] Le forze dell'ordine straniere hanno utilizzato l'OSINT in diversi casi, tra cui le indagini su un attacco terroristico e una rapina a mano armata, oltre che per la ricerca di una persona scomparsa, la caccia a un molestatore sessuale e la preparazione di un personaggio sotto copertura.[61]

3.5.4 Penetration Testing

Una fase importante nella preparazione di un test di penetrazione è la raccolta dei dati. A seconda degli obiettivi definiti, ciò può essere facilitato dall'uso di strumenti e servizi che sfruttano le informazioni pubbliche, ad esempio l'enumerazione dei sottodomini di un sito web utilizzando strumenti come Sublist3r o motori di ricerca specializzati come Spyse o Shodan.

3.5.5 Ingegneria sociale e human intelligence

Il celebre hacker Kevin Mitnick, pioniere dell'ingegneria sociale descrive il primo passo di ogni attacco di quest'ultima come la raccolta e la valutazione di informazioni da fonti pubbliche disponibili.[62] Lekati descrive anche la combinazione di OSINT, SOCMINT e HUMINT: "OSINT e SOCMINT possono essere utilizzate come discipline di supporto quando l'obiettivo finale di un investigatore o di un professionista dell'intelligence è quello di essere in grado di interagire efficacemente con un sospetto e di infiltrarsi in un gruppo, di reclutare l'obiettivo, di ottenere una confessione o di condurre altre attività principalmente legate alla HUMINT".[63] L'OSINT e la SOCMINT presentano ampie sovrapposizioni; la SOCMINT è emersa come una sottodisciplina dell'OSINT e continua a essere ampiamente suddivisa nell'ambito dell'Open Source Intelligence.

3.5.6 Rintracciamento pubblico

L'agenzia di polizia Europol ha istituito il progetto "Stop Child Abuse" in cui il pubblico viene interrogato per identificare i dettagli delle immagini utilizzando pezzi estratti da materiale sessualmente esplicito che coinvolge minori.[64]

3.5.7 Ricerca e salvataggio di persone scomparse

Dopo la misteriosa scomparsa del vincitore del premio Turing Jim Gray durante una gita in barca nel 2007,[65] è stata avviata un'esercitazione civile di ricerca e salvataggio senza precedenti che prevedeva l'analisi automatizzata e in *crowdsourcing* di immagini satellitari e viste aeree dell'area in cui era scomparso.[66]

3.5.8 Protezione civile

Un'applicazione molto simile si trova all'interno delle unità di protezione civile. "Dopo il devastante terremoto di Haiti del 2010, sono state fondate le [Volunteer & Technical Communities] V&TC con l'obiettivo di elaborare e fornire dati pubblicamente disponibili alle forze di emergenza e alla popolazione"[67]. Nel 2016 è stato fondato il Virtual Operations Support Team (VOST) dall'Agenzia federale tedesca per i soccorsi tecnici in collaborazione e con il supporto dei ricercatori dell'Università di Wuppertal. "In situazioni operative, il gruppo 'Digital Situation Exploration' ottiene informazioni dai social media e da altre fonti pubbliche utilizzando metodi di valutazione dei Big Data per elaborarle e presentarle in modo facile da usare. Ciò include l'identificazione di informazioni errate e di disinformazione, ma anche la verifica e la geo-localizzazione delle informazioni rilevanti per la situazione"[67].

3.5.9 Gestione del rischio informatico

Il monitoraggio e l'analisi delle fonti pubbliche possono supportare un'efficiente valutazione del rischio. Questa può essere supportata da strumenti adattati ai requisiti specifici della rispettiva organizzazione.[68] Durante le valutazioni del rischio, spesso vengono utilizzati servizi per stimare l'attuale livello di sicurezza. Fornitori di servizi come BlackKite "[utilizzano] Open-Source Intelligence (OSINT) e scansioni informatiche non intrusive per identificare potenziali rischi per la sicurezza".[69]

3.5.10 Preparazione di un atto criminale

Il personaggio televisivo Kim Kardashian ha fatto notizia nel 2016 quando è stata rapinata a mano armata nella sua stanza d'albergo a Parigi e derubata dei suoi gioielli per un valore di sei milioni di euro.[70] Quando uno dei sospetti è stato catturato, ha confessato alla polizia che lui e i suoi complici hanno analizzato i post di Instagram e altre fonti Internet in combinazione con le informazioni di una persona vicina a Kardashian per pianificare e commettere il crimine.[71]

4 Metodi e tools OSINT

4.1 Sfide dell'utilizzo di strumenti commerciali off-the-shelf

L'Intelligence Community utilizza generalmente strumenti commerciali off-the-shelf (COTS)[4] per l'analisi OSINT, in particolare per l'analisi dei dati dei social media. La maggior parte degli strumenti COTS è stata sviluppata per scopi commerciali: pubblicità, gestione del marchio e analisi dei consumatori. Le aziende vogliono capire e prevedere il comportamento di acquisto dei clienti, posizionare i loro prodotti in modo che siano disponibili quando il cliente è più influenzabile e influenzare l'opinione del cliente sul prodotto o sull'azienda stessa. Questi strumenti possono spesso servire gli interessi della IC, ma in quel contesto il loro utilizzo può risultare limitato perché non sono stati progettati per i suoi scopi.

Inoltre, il panorama di questi strumenti è estremamente estremamente attivo e soggetto a cambiamenti. Questo problema si manifesta in vari modi:

- I feed di dati possono essere limitati o eliminati dall'azienda che possiede i contenuti, per una serie di motivi
- Le aziende possono voler proteggere i dati degli utenti o, al contrario, possono iniziare a vendere dati che prima erano disponibili gratuitamente.
- Le aziende potrebbero aver acquisito una capacità o averne sviluppata una propria per l'analisi dei contenuti dei social media e potrebbero voler minare le capacità concorrenti eliminando la loro fonte di dati.

Ad esempio, Topsy era un servizio di analisi dei social media che indicizzava tutti i tweet pubblicati su Twitter e forniva funzioni di ricerca gratuite. Dopo otto anni, il servizio è stato inaspettatamente disattivato il 15 dicembre 2015[72], due anni dopo essere stato acquisito da Apple.[73] Questo caso è esemplificativo per i servizi di analisi e le operazioni della IC che si affidano ad altri servizi per le prime fasi del ciclo di acquisizione e analisi dei dati.

La IC è abituata ad accessi ai dati inaspettatamente non disponibili. Gli analisti che "raccolgono" dati nell'ambito della SIGINT possono perdere l'accesso per una serie di motivi, tra cui riconfigurazioni del sistema e nuove crittografie. Gli analisti che "raccolgono" dati relativi alla HUMINT sono alle prese con la possibilità di compromettere una fonte. I malfunzionamenti dei satelliti possono lasciare i raccoglitori IMINT al buio. I consumatori di intelligence possono essere frustrati dalla perdita di un flusso di informazioni raccolte con metodi segreti, ma la perdita può essere spiegata come conseguenza inevitabile dei metodi non aperti: la fonte di dati non è più accessibile per essere analizzata. Un vantaggio dell'OSINT è che è più affidabile dei metodi di raccolta occulta. L'improvvisa perdita di un feed di dati open-source - quando i dati grezzi sono ancora accessibili online - può essere ingiustamente interpretata come un riflesso negativo sull'OSINT da parte dei consumatori di intelligence che potrebbero non essere né consapevoli né interessati al processo di trasformazione dei dati grezzi in un prodotto di intelligence. Quando Twitter è ancora online e le persone continuano a twittare, può essere più difficile spiegare a un consumatore di intelligence perché un prodotto OSINT non è più improvvisamente disponibile.

Affidandosi a strumenti COTS, l'IC rischia di rimanere sempre indietro rispetto allo stato dell'arte dell'analisi dei social media a causa del tempo necessario per completare le verifiche necessarie all'utilizzo. La predominanza di startup in questo settore complica la capacità dell'IC di costruire un rapporto di fiducia con fornitori affermati per semplificare il processo di verifica. L'IC potrebbe naturalmente sviluppare strumenti interni, ma si tratta di un'alternativa costosa. L'analisi dei social media è un mercato dinamico anche grazie ai rapidi miglioramenti della potenza di calcolo e delle capacità di elaborazione dei dati. Gli strumenti sono

[4] L'espressione "componente COTS" o componente OTS, in inglese (Commercial) Off-the-Shelf component, si riferisce a componenti hardware e software disponibili sul mercato per l'acquisto da parte di aziende di sviluppo interessate a utilizzarli nei loro progetti.

sempre più capaci di gestire grandi quantità di dati e l'apprendimento automatico sta facendo passi da gigante. Invece di dover insegnare ai computer come svolgere attività complesse, si stanno costruendo sistemi che permettono ai computer di imparare a svolgerle da soli.[74].

4.2 Metodi utilizzati nell'analisi dei contenuti dei social media

Nonostante gli strumenti di raccolta OSINT si evolvano quasi quotidianamente, i vari metodi utilizzati dagli strumenti cambiano meno drasticamente.

La maggior parte degli strumenti utilizzano metodi esistenti già da prima della loro applicazione ai contenuti su Internet, ma la vasta proliferazione di piattaforme di social media e la sempre maggiore facilità con cui gli individui possono accedere a Internet rendono questo ambiente ricco per la raccolta di informazioni di intelligence. Questi sono:

- analisi lessicale
- analisi dei social network
- analisi geospaziale

Le combinazioni di questi metodi vengono utilizzate per isolare, descrivere e analizzare i dati. Distinguere tra gli strumenti analitici open-source disponibili può essere difficile, a causa della loro abbondanza e delle scarse descrizioni. L'identificazione delle componenti specifiche dei metodi che utilizzano, tuttavia, fornisce un criterio per valutare e confrontare le capacità.

Gli strumenti possono essere confrontati in termini di quantità di metodi analitici che possono utilizzare e di velocità, accuratezza e capacità di eseguire analisi.

4.3 Analisi dei Social network

Tabella degli strumenti testati sull'analisi dei social network

Per decenni, prima dell'avvento della più recente generazione di applicazioni basate sul web, l'analisi delle reti sociali ha cercato di spiegare le relazioni tra gli individui come una serie di scambi che possono essere mappati e tracciati per spiegare le interazioni passate e prevedere quelle future. I principi alla base dell'analisi delle reti sociali sono i seguenti:[75]

- Gli attori sono visti come interdipendenti, non autonomi.
- I legami relazionali tra gli attori sono canali per il trasferimento o il "flusso" di risorse (materiali e non).
- I modelli di rete considerano l'ambiente strutturale come un'opportunità o un vincolo per l'azione individuale.
- I modelli di rete concettualizzano la struttura (sociale, economica, politica, ecc.) come ultimo modello di relazioni tra gli attori. di relazioni tra gli attori.

L'analisi delle reti sociali nell'era di Internet ha creato un'offerta sovrabbondante di nuovi dati per lo studio delle interazioni in rete, mentre i nuovi strumenti dei social media offrono una maggiore visibilità delle reti.

Gli elementi fondamentali dell'analisi delle reti sociali sono presentati nella Figura 5. Ogni unità di una rete sociale è descritta come un nodo. I nodi possono essere individui esterni o interni a una rete, ma l'analisi delle reti sociali si concentra principalmente sui nodi che fanno parte di gruppi più ampi. Una diade è costituita da due nodi che interagiscono tra loro, come indicato dalla linea che collega A a B. Una triade è, allo stesso modo, un'interazione tra tre nodi - A, B e C. A partire da questi elementi di base, si formano reti più ampie in

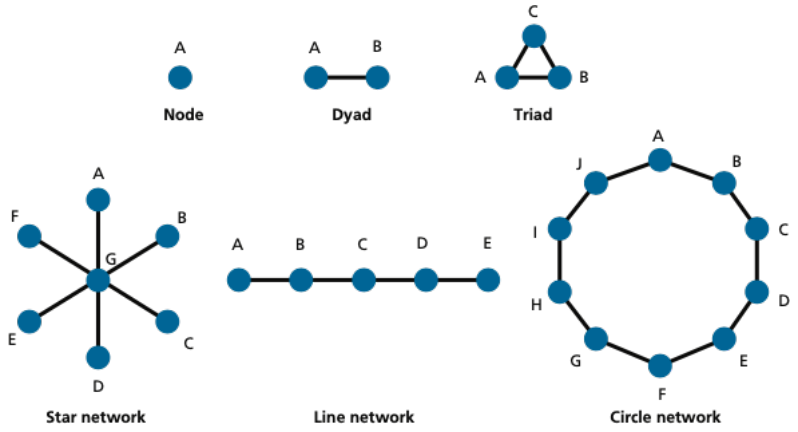


Figura 5: Diagramma di analisi dei social network[3]

grado di descrivere i modi in cui i nodi interagiscono tra loro, quali nodi detengono un maggiore controllo o potere e come i nodi sono collegati tra loro attraverso connessioni condivise. La rete a stella, la rete a linee e la rete a cerchi sono modi per visualizzare diversi tipi di interazioni, descritti con esempi di seguito.[76]

4.3.1 Grado

Il grado è il numero di connessioni di un nodo; più grande è il grado, più connessioni ha il nodo. Nella Figura 5, il grado è illustrato dalla posizione del nodo G nella rete a stella: G ha grado sei, mentre tutti gli altri nodi hanno grado uno, il che significa che G avrà maggiori opportunità di accesso alle informazioni o maggiore capacità di influenza rispetto a tutti gli altri nodi della rete.

4.3.2 Densità

In qualsiasi grafo esiste un numero finito di linee e il numero di nodi ne determina il massimo. La densità è il rapporto tra le linee effettivamente presenti nel grafo e il massimo teorico possibile. Nella Figura 5, la rete circolare ha una bassa densità, il che significa che c'è un basso numero di linee tra i nodi rispetto al numero che potrebbe esistere (ad esempio, A e F potrebbero essere collegati, come anche I e C). Maggiore è questo rapporto, più interazioni si verificano all'interno di un gruppo, il che può essere misurato come coesività. All'interno di gruppi più grandi di due persone, questo indica "la misura in cui i membri della rete si conoscono e interagiscono tra loro ".[77]

4.3.3 Betweenness

La *betweenness* è un'indicazione del grado di controllo della comunicazione da parte di un singolo punto (o nodo).[78] Nella rete lineare della Figura 5, B si trova tra A e C; pertanto, qualsiasi informazione che A voglia trasmettere a C deve passare attraverso B, il che significa che B può controllare il messaggio che C riceve, cambiandolo o impedendo che raggiunga completamente C. L'analisi delle reti sociali può utilizzare misure di betweenness per designare gli individui come "influencer" all'interno di una data rete, vedendo come il linguaggio cambia e si trasforma nello scambio da un attore all'altro, come in una gigantesco gioco di telefono senza fili.

Centralità della *"betweenness"* La centralità della betweenness suggerisce un modo per determinare come reti o individui altrimenti non associati condividano un legame comune che consente la comunicazione tra loro. Quando due reti diverse interagiscono, ad esempio, se la rete a stella e la rete circolare avessero una connessione condivisa, le misure della *"betweenness centrality"* indicano come questi due gruppi siano collegati e forniscono informazioni sull' "eterogeneità o variabilità della betweenness nell'intero insieme di attori".[79]

4.3.4 Vicinanza

Mentre la betweenness indica quali individui all'interno di un dato gruppo potrebbero controllare il messaggio, la vicinanza misura quanto ogni individuo di un gruppo sia indipendente o dipendente dagli altri e quindi quanto una persona dipenda da un'altra per trasmettere un messaggio.[78] Nella rete lineare della Figura 5, l'attore C è più vicino a tutti gli altri attori della rete, mentre gli attori A ed E sono più lontani.

4.3.5 Misure di centralità

Le misure di centralità descrivono l'importanza di un singolo nodo all'interno di una rete più ampia. Gli individui con alta centralità hanno "un elevato coinvolgimento in molte relazioni, indipendentemente dalla direzionalità di invio/ricezione o dal volume di attività".[80] Nel contesto di un'interazione su Twitter, un utente con alta centralità verrebbe frequentemente menzionato da altri utenti, indipendentemente dal fatto che l'utente abbia iniziato o meno la conversazione, e inizierebbe anche frequentemente interazioni con altre persone nella sua rete.

4.3.6 Direzionalità

Misurata come "grado in uscita" (cioè le informazioni che escono) o "grado in entrata" (cioè le informazioni che entrano), gli attori con grado in entrata maggiori tendono a essere i più prestigiosi o importanti all'interno di una rete.[80] La direzionalità, a differenza delle altre misure qui definite, guarda dove le informazioni hanno avuto origine e in quale direzione fluiscono. Nella rete lineare, ad esempio, l'importanza di un nodo dipende dalla direzione in cui fluiscono le informazioni; se tutte le informazioni fluissero verso destra, il nodo E avrebbe il grado in entrata maggiore. Nelle diadi, la direzionalità può indicare se entrambi i membri condividono la stessa influenza/potere nell'interazione o se c'è una dinamica di potere ineguale.[77]

4.3.7 Applicare gli strumenti di analisi nei social media

Twitter è probabilmente il più importante strumento di social media utilizzato dagli analisti delle reti sociali per illustrare e studiare i principi classici dell'analisi delle reti. Twitter fornisce agli utenti degli username unici, o "handle", che poi usano per interagire direttamente con altri utenti rispondendo a una conversazione disponibile pubblicamente o iniziando una conversazione con un altro utente noto, un processo chiamato "menzione" in un thread di conversazione. Tutte queste interazioni sono pubbliche. Questo permette agli osservatori di vedere chi interagisce con chi, chi è connesso a chi e attraverso chi, e la qualità e la quantità di queste interazioni.

Un esempio di come l'analisi delle reti sociali sia attualmente impiegata per comprendere un problema, e possibilmente influenzarne l'evoluzione, è il monitoraggio del propagarsi di ideologie estremiste online. Ad esempio, uno studio RAND del 2016 ha utilizzato i dati dei social media e l'analisi lessicale per confrontare i sostenitori dell'ISIS con i suoi detrattori e determinare gli schemi che offrivano opportunità di influenzare queste comunità.[81] Seguendo le parole chiave, o "hashtag", è possibile seguire lo sviluppo di una nuova parola chiave e osservarne la diffusione, prima tra i singoli utenti, poi in reti più ampie e quindi attraverso le reti, eventualmente assumendo un nuovo significato o una nuova definizione a ogni nuova condivisione. Una volta compresi i modelli di comunicazione e il linguaggio condiviso di una comunità, potrebbe essere possibile

condurre campagne di contro-messaggistica che utilizzino le stesse parole chiave usate dagli account pro-ISIS per mostrare le conseguenze negative del coinvolgimento nell'organizzazione.

4.3.8 Strumenti per l'analisi dei Social Network

Accountanalysis Questo strumento consente di esaminare gli account Twitter. Ad esempio, quanto sono automatizzati, quanti Retweet pubblicano o a quali siti web si collegano più spesso.

La Scheda utente mostra le informazioni di base dell'account che si sta analizzando. Quanti tweet ha fatto, quanti account segue e quanti li seguono. Il numero di tweet preferiti e il numero di liste in cui sono presenti. La maggior parte di queste metriche può essere modificata direttamente o indirettamente dall'utente. Solo la data di creazione e l'ID sono fissi.

Il grafico Ritmo giornaliero mostra in quali ore e in quali giorni della settimana un account è più attivo. Più il blu è scuro, più tweet sono stati pubblicati in quel momento. È possibile fare clic su ogni singola ora o selezionare tutte le ore di un giorno feriale o la stessa ora di ogni giorno feriale facendo clic sulla rispettiva etichetta. O qualsiasi combinazione. Gli orari sono indicati nel fuso orario locale. Con questo grafico è possibile notare se l'account target è un bot se non è presente un pattern che indichi delle ore di sonno da parte di una persona.

Il volume dei tweet per data mostra il numero di tweet pubblicati dall'account in una data specifica. È possibile fare clic e trascinare per visualizzare solo i Tweet pubblicati nelle date evidenziate. Sono poche le persone che pubblicano lo stesso numero di tweet ogni giorno. Singoli giorni con un volume elevato possono indicare che si è verificato un evento. Il grafico degli hashtag può aiutare a identificare il tipo di evento.

Il grafico del giorno della settimana mostra in quali giorni della settimana un account è più attivo. È possibile fare clic su ciascun giorno della settimana per selezionare i tweet pubblicati in quel giorno. Il grafico mostra il numero aggregato di tweet per quel giorno della settimana, provenienti da tutti i tweet recuperati dall'account.

Twitter cerca di determinare in quale lingua è scritto ogni Tweet. Il grafico Lingua dei tweet mostra queste lingue e quanti tweet l'account ha pubblicato in esse. L'algoritmo di categorizzazione non è perfetto. Soprattutto i Tweet brevi o quelli con slang danno luogo a false categorizzazioni. I tweet con un'elevata incertezza (ad esempio i tweet di soli media) sono classificati come "Sconosciuti".

Il grafico dell'interfaccia utilizzata mostra quali app ha usato l'account per pubblicare i tweet. La maggior parte degli utenti utilizza quelle ufficiali (Twitter Web App, Twitter per Android, Twitter per iPhone, Twitter per iPad, TweetDeck), ma ci sono anche alcune applicazioni di terze parti molto diffuse e complete (Tweetbot, Fenix,...). Altre app di terze parti sono utilizzate soprattutto per l'assistenza ai clienti (swat.io, hootsuite,...) o per l'automazione (Buffer, IFTTT,...).

Ogni Tweet può contenere uno o più Hashtag. La sezione Hashtag usati li mostra tutti e consente di filtrare i Tweet che contengono gli Hashtag di interesse. Gli hashtag hanno molteplici usi. La maggior parte degli account li usa per categorizzare i propri Tweet, partecipare a eventi e/o aumentare la visibilità dei Tweet. Osservare gli hashtag più utilizzati aiuta a capire quali sono gli argomenti principali su cui un account twitta.

Ogni Tweet può contenere uno o più URL. Il grafico dei nomi host degli URL mostra i nomi host di tali URL. Il nome host contiene il sottodominio ma non il percorso, la query o il frammento dell'URL. "twitter.com" e "mobile.twitter.com" sono esclusi in quanto rappresentano Tweet e caricamenti multimediali citati.

Il grafico Utenti rispondenti mostra gli account a cui l'account analizzato scrive il maggior numero di risposte. Vengono conteggiati solo gli utenti a cui l'account ha risposto direttamente. Se l'utente A pubblica un Tweet, l'utente B risponde e l'utente C risponde a tale risposta, solo l'utente B compare nel grafico quando si analizza l'utente C. Le auto-risposte sono escluse.

Il grafico Utenti "retwittati" mostra gli account che l'account analizzato retwitta più spesso. Il grafico Utenti citati mostra gli account che l'account analizzato cita più spesso. La griglia Ultimi tweet visualizza i tweet della selezione corrente. Mentre si scava nei dati, è utile guardare i tweet veri e propri. L'etichetta a

destra mostra il tipo di Tweet specifico. Facendo clic sulla data è possibile visitare il Tweet su Twitter. Facendo clic su "Mostra l'intero Tweet" è possibile visualizzare l'immagine Twitter del Tweet.

Inizialmente non carica tutti i tweet dell'account target, ma solo i più recenti. Per caricare tutti i tweet è necessario l'account pro a pagamento.

Digital Image Forensic Analyzer Digital Image Forensic Analyzer è un sito web che fornisce alcuni servizi (già presenti in Ghiro) di analisi forense su immagini, ma con l'obiettivo di offrire un servizio stabile e affidabile per gli investigatori forensi e i professionisti della sicurezza.

Facendo l'upload di un'immagine il sito mostra una serie di informazioni estratte quali:

- Analisi statica
- Estrazione dei metadati EXIF
- Estrazione dei metadati IPTC
- Estrazione di metadati XMP
- Estrazione dell'anteprima dai metadati
- Geolocalizzazione
- Error Level Analysis (ELA)
- Controllo della firma

Secondo quanto dichiarato dal sito, l'applicazione è stata progettata con i seguenti obiettivi:

- Fornire le principali tecniche di image forensics in un'unica applicazione.
- Privacy: Le foto e i report caricati sono privati. Il report è accessibile solo se si conosce il suo link diretto. Le immagini e le analisi vengono memorizzate per 7 giorni, dopodiché è necessario inviarle nuovamente, quindi il link per un'analisi scade dopo 7 giorni.
- Gratuità.
- Non divulgazione: i dati di analisi sono strettamente privati. Nessuna condivisione dei dati, nessuna condivisione dell'analisi.

Viene inoltre fornita una semplice API REST per l'utilizzo automatico. È possibile sviluppare il proprio strumento di invio automatico su di essa.

Ghiro Si tratta di uno strumento completamente automatizzato, progettato per eseguire analisi forensi su grandi quantità di immagini, utilizzando un'applicazione web di facile utilizzo.

È possibile controllare tutte le funzioni di Ghiro tramite l'interfaccia web, caricare un'immagine o un gruppo di immagini per ottenere una panoramica rapida e approfondita dell'analisi delle immagini e raggruppare le immagini in casi e cercare qualsiasi tipo di dati di analisi.

Le caratteristiche principali di Ghiro sono:

- Estrazione dei metadati: I metadati sono suddivisi in diverse categorie a seconda dello standard da cui provengono, i metadati delle immagini vengono estratti e classificati. EXIF, IPTC, XMP.
- Localizzazione GPS: Incorporata nei metadati dell'immagine, a volte c'è un geotag, dati GPS che fornisce la longitudine e la latitudine del luogo in cui è stata scattata la foto, viene letto e la posizione viene visualizzata sulla mappa.
- Informazioni MIME: Il tipo MIME dell'immagine rilevato per conoscere il tipo di immagine con cui abbiamo a che fare, sia in forma contattata che estesa.

- ELA: ELA sta per Error Level Analysis. Identifica le aree all'interno di un'immagine che si trovano a livelli di compressione diversi. L'intera immagine dovrebbe avere all'incirca lo stesso livello, se viene rilevata una differenza, probabilmente indica una modifica digitale.
- Estrazione delle miniature: Le miniature e i relativi dati vengono estratti dai metadati dell'immagine e archiviati per la revisione.
- Coerenza delle miniature: A volte, quando una foto viene modificata, l'immagine originale viene modificata ma le miniature non presentano differenze tra le miniature e le immagini.
- Signature Engine: dispone di oltre 120 firme che forniscono prove sui dati più critici per evidenziare i punti focali e le esposizioni comuni.
- Hash Matching: Supponiamo di cercare un'immagine e di avere solo il valore hash. Possiamo fornire un elenco di hash e tutte le immagini corrispondenti vengono segnalate.

GvngSearch GvngSearch è uno strumento che consente la raccolta di dati tramite OSINT. Raccoglie vari strumenti osint in un solo programma. Alcune scritte sono in spagnolo. Dopo averlo avviato si può scegliere se focalizzare la ricerca su un sito web o su un singolo individuo.

Nella sezione "Dox personal" scegliendo "Nickname research" si può ricercare la presenza di un nickname in una selezione ristretta di siti, in maniera molto simile a holehe, non sembrano però risultare falsi positivi.

Scegliendo "Validate email" si viene invitati a inserire un indirizzo email. Inserendo il mio personale vengo invitato a inserirlo correttamente.

Inserendo un indirizzo ip (Anche di un sito web) sulla terza opzione "geolocate by IP" mostra una serie di info geografiche tra cui latitudine, longitudine, città, ISP e relativo link di Google.

Quando si seleziona "Go back" il programma crasha.

La sezione "Dox web" permette di scegliere diverse singole operazioni quali Traceroute, Ping Test, DNS Lookup, Reverse DNS, Find DNS Host, Find Shared DNS, Zone Transfer, Whois Lookup, IP Location Lookup, Reverse IP Lookup, TCP Port Scan, Subnet Lookup, HTTP Header Check, Extract Page Links. Abbastanza scomodo poiché ogni volta bisogna scegliere l'operazione e inserire il target.

In generale un pacchetto di strumenti con funzionalità già presenti e abbastanza raffazzonato.

Have I been Pwnd? Have I Been Pwned? (HIBP) è un sito web che consente agli utenti di Internet di verificare se i propri dati personali sono stati compromessi da violazioni dei dati. Il servizio raccoglie e analizza centinaia di dump e pastebin di database contenenti informazioni su miliardi di account violati e consente agli utenti di cercare le proprie informazioni inserendo il proprio nome utente o indirizzo e-mail. Gli utenti possono anche registrarsi per ricevere una notifica se il loro indirizzo e-mail appare nei dump futuri. Il sito è stato ampiamente pubblicizzato come una risorsa preziosa per gli utenti di Internet che desiderano proteggere la propria sicurezza e privacy. Il sito è stato creato dall'esperto di sicurezza Troy Hunt il 4 dicembre 2013.

A giugno 2019 Have I Been Pwned? conta in media circa centosessantamila visitatori giornalieri, il sito ha quasi tre milioni di abbonati e-mail attivi e contiene record di quasi otto miliardi di account.

Nel settembre 2014, Hunt ha aggiunto una funzionalità che ha consentito di aggiungere automaticamente nuove violazioni dei dati al database di HIBP. La nuova funzionalità ha utilizzato Dump Monitor, un bot di Twitter che rileva e trasmette i probabili dump delle password trovati sui pastebin, per aggiungere automaticamente nuove potenziali violazioni in tempo reale. Le violazioni dei dati spesso compaiono sui pastebin prima che vengano segnalate a tutti; quindi, il monitoraggio di questa fonte consente ai consumatori di essere informati prima se sono stati compromessi.

- Funzione principale - La funzione principale di Have I Been Pwned? da quando è stato lanciato è quella di fornire al grande pubblico un mezzo per verificare se le loro informazioni private sono divulgate o compromesse. I visitatori del sito web possono inserire un indirizzo e-mail e visualizzare un elenco di tutte le violazioni dei dati note con i record legati a tale indirizzo e-mail. Il sito web fornisce anche dettagli su ciascuna violazione dei dati, come il retroscena della violazione e quali tipi specifici di dati sono stati inclusi in essa.

- Notify Me - Have I Been Pwned? offre anche un servizio "Notify Me" ("Avvisami") che consente ai visitatori di iscriversi al sito e ricevere le notifiche su future violazioni. Una volta che qualcuno si iscrive a questo servizio riceverà un messaggio di posta elettronica ogni volta che le sue informazioni personali vengono rilevate in una nuova violazione dei dati.

- Ricerca per dominio - La ricerca per dominio consente di trovare tutti gli indirizzi e-mail di un determinato dominio che sono stati coinvolti in una delle violazioni dei dati attualmente presenti nel sistema. È inoltre possibile ricevere notifiche se l'indirizzo compare in future violazioni, fornendo un'e-mail di notifica. Prima di poter eseguire una ricerca di dominio, è necessario verificare di avere il controllo del dominio che si sta cercando. Se non è possibile verificare il controllo del dominio, non sarà possibile cercare gli indirizzi e-mail violati su di esso.

- Who's been pwnd - Questa sezione mostra una panoramica delle varie violazioni che sono state consolidate in Have I Been Pwned. Queste sono accessibili programmaticamente tramite l'API HIBP e anche tramite il feed RSS. Ogni entry dà una breve descrizione della violazione e dati come data della violazione, data dell'aggiunta a HIBP, numero degli account compromessi e il tipo di dati compromessi

- Password - Nell'agosto 2017, Hunt ha reso pubbliche 306 milioni di password a cui era possibile accedere tramite una ricerca sul Web o scaricabili in blocco.

  Nel febbraio 2018, l'informatico britannico Junade Ali ha creato un protocollo di comunicazione (utilizzando il k-anonimato e la funzione crittografica di hash) per verificare in modo anonimo se una password fosse trapelata senza rivelare completamente la password cercata. Questo protocollo è stato implementato come API pubblica nel servizio di Hunt ed è ora utilizzato da più siti web e servizi, inclusi i gestori di password e le estensioni del browser. Questo approccio è stato successivamente replicato dalla funzione di controllo password di Google. Ali ha lavorato con gli accademici della Cornell University per analizzare formalmente il protocollo per identificare i limiti e sviluppare due nuove versioni di questo protocollo note come Frequency Size Bucketization e Identifier Based Bucketization. Nel marzo 2020, il pudding crittografico è stato aggiunto a questo protocollo.

Holehe Holehe controlla se un'e-mail è collegata a un account su siti come Twitter, Instagram, Imgur e oltre 120 altri.

L'installazione e la distribuzione di Holehe all'interno di un ambiente con interfaccia a riga di comando è molto semplice, il che rende questo script Python uno dei più facili da usare. Lo script funziona prendendo l'indirizzo e-mail specificato dall'utente e verificandolo attraverso la funzione "password dimenticata". L'obiettivo non viene avvisato di questa azione; ad esempio, non riceverà un'e-mail di ripristino della password. I social media più popolari, come Facebook, non sono inclusi nell'elenco delle fonti online di Holehe perché una richiesta di "password persa" su Facebook fa scattare naturalmente un avviso all'utente. Per quanto riguarda le altre pagine web, è probabile che esse dispongano di misure di privacy rafforzate, il che significa che una richiesta di "password smarrita" non verificherà un indirizzo e-mail. Ad esempio, se si invia una richiesta di "password smarrita" a codecanyon.com, il sito web risponderà con qualcosa del tipo "Se è stato trovato un account corrispondente, è stata inviata un'e-mail a user@email.com per consentire di reimpostare la

password". In questo caso, ciò non è ideale per Holehe - o per noi, gli Investigatori Digitali - poiché la risposta del sito web non conferma né smentisce l'esistenza di un account registrato al nostro indirizzo e-mail.

Lo strumento è stato ulteriormente sviluppato per diventare una trasformazione di Maltego, il che è certamente ideale per gli investigatori digitali che utilizzano Maltego. L'utilizzo di questo strumento presenta alcuni aspetti negativi, nessuno dei quali è imputabile agli sviluppatori. Ad esempio, molte pagine web hanno implementato meccanismi di limitazione della velocità che impediscono a script come Holehe di ottenere i dati dell'utente. Per risolvere questo problema, gli sviluppatori suggeriscono agli utenti di utilizzare semplicemente una VPN e di cambiare l'indirizzo IP durante ogni ciclo di scansione. Un altro aspetto negativo è la prospettiva che la maggior parte delle pagine web implementino salvaguardie aggiuntive per la reimpostazione delle password, come nell'esempio indicato nel paragrafo precedente. A parte questo, Holehe è un ottimo strumento, è stato ben realizzato e fornisce i risultati che gli investigatori digitali si aspettano. È facile da installare, veloce da usare e molto potente in termini di risultati.

Ignorant Tool molto semplice in grado di controllare qualora un numero telefonico in input sia collegato ad un account Amazon, Instagram e Snapchat.

Mosint Mosint è uno strumento gratuito e open-source scritto in Python disponibile su GitHub. Dispone di diversi moduli che eseguono diverse operazioni per effettuare la ricognizione degli indirizzi e-mail. Tra questi vi sono:

- Verifica dell'email target
- Scansione sociale dell'e-mail target
- Trova email e domini correlati con l'email target
- Trova le leaks di password per l'e-mail target
- Ricerca dei dump pastebin per l'email target
- Trova ulteriori informazioni sul dominio dell'email target
- Output dei risultati su file di testo

Mr. Holmes Mr.Holmes è uno strumento di raccolta di informazioni il cui scopo principale è quello di ottenere informazioni su domini, nomi utente e numeri di telefono con l'aiuto di fonti pubbliche disponibili su Internet e di utilizzare l'attacco google dorks per ricercatori specifici. Utilizza anche dei proxy per rendere le richieste completamente anonime e un WhoIS Api per ottenere ulteriori informazioni su un dominio.

Possibile impostare la lingua del tool in italiano. Il tool è composto da vari moduli:

- Social account: Ricerca su vari siti e social network le occorrenze di account che hanno come username l'input dell'utente, molto simile a ciò che fa il tool Sherlock. (È possibile impostare un proxy per questa ricerca). Una volta trovati gli account è possibile effettuare uno scraping dei dati di quest'ultimi. Come ultima ricerca è possibile effettuare una ricerca automatizzata google dork.
- Numero di telefono: Ricerca quante più informazioni possibili sul numero telefonico in input comprese geolocalizzazione, google dorks e info da siti di lookup.
- Domain/ip: Ricerca quante più informazioni possibili sul dominio in input, facendo ricerca WhoIs, Reputation Scan, port scan e traceroute.

- Database: con questo comando è possibile avviare in localhost una gui per visualizzare i risultati delle ricerche già effettuate.
- Port scanner: È possibile inserire l'indirizzo ip di un host e impostare un intervallo di porte da scannerizzare, scannerizzare tutte le porte disponibili o solo delle porte specifiche .
- Email: controlla se la email è valida, se è presente nel database di haveibeenpwnd o di intelligence X.

Octosuite Octosuite è un strumento OSINT che permette di ottenere informazioni su account GitHub. Dato l'username di un account presente su Github, il programma sarà in grado di trovare:

- Informazioni sul profilo di un'organizzazione
- Gli eventi di un'organizzazione
- Le repository di un'organizzazione
- I membri pubblici di un'organizzazione
- Le informazioni di una repository
- I collaboratori di una repository
- Gli stargazer di una repository
- I fork di una repository
- Le release di una repository
- L'elenco di file in un percorso specifico di una repository
- Le informazioni sul profilo di un utente
- I gist di un utente

Osintgram Osintgram è un programma in grado di ricavare diverse informazioni da un account Instagram. È necessario disporre di un account per accedere a queste informazioni, le credenziali di esso vanno inserite nel file credentials.ini per fare in modo che le query vadano a buon fine. È possibile esportare i risultati delle ricerche in file json o txt. Le informazioni ottenibili sono:

- tutti gli indirizzi registrati dalle foto del target
- didascalie delle foto del target
- elenco di tutti i commenti ai post del target
- totale dei commenti ai post del target
- followers del target
- utenti seguiti dal target
- email dei followers del target
- email degli utenti seguiti dal target
- numero di telefono dei follower del target
- numero di telefono degli utenti seguiti dal target
- hashtag utilizzati dal target
- info sul target
- totale dei like ai post del target
- tipo di post del target (foto o video)
- descrizione delle foto del target
- foto del target nella cartella di output
- immagine del profilo del target
- storie del target
- elenco degli utenti taggati dal target
- elenco degli utenti che hanno commentato le foto del target
- elenco degli utenti che hanno taggato il target

PimEyes PimEyes è un sito che permette di eseguire ricerche inversi di immagini basate sui volti. Il sito permette di inserire immagini e scegliere quale tra i volti rilevati nella foto deve essere oggetto di ricerca,

aggiungere altre foto che comprendono lo stesso volto al fine di affinare la ricerca. È inoltre possibile specificare un intervallo di date entro le quali le foto.

recon-ng Recon-ng è un software gratuito e open source scritto da Tim Tones, completamente in Python e disponibile su GitHub; è composto da moduli indipendenti, aiuti interattivi e molteplici plugin integrati. L'interfaccia di Recon-ng a riga di comando è molto simile a quella di Metasploit 1 e Metasploit 2. Esso serve ad effettuare tecniche di ricognizione, ossia la fase direttamente connessa al Footprinting, nella quale si ricavano informazioni dal target al fine di trovare vulnerabilità e poterle sfruttare in seguito.

I moduli da cui è composto sono divisi in macro categorie, che sono:

- recon: moduli di ricognizione;
- reporting: moduli per l'output delle informazioni ottenute (in file csv, html, etc.);
- discovery;
- import: per importare una lista di host o file csv;
- exploit: per effettuare exploit (ma non è stato creato per questo fine, infatti ne ha solo due).

Sherloq Sherloq è un programma che offre una raccolta di strumenti open source per l'analisi forense di immagini digitali. Una volta avviato il programma si mostra come completo di gui. In alto a sinistra il pulsante "Load image" permette di caricare un'immagine da analizzare. Nella parte sinistra della finestra si possono scegliere i vari tipi di analisi alle quali sottoporre le immagini. Esse sono: Generale Immagine originale: visualizza l'immagine di riferimento inalterata per l'ispezione visiva. File Digest: recupera le informazioni sui file fisici e gli hash crittografici e percettivi. Editor esadecimale: apre un editor esadecimale esterno per mostrare e modificare i byte grezzi. Ricerca simile: sfoglia i servizi di ricerca online per trovare immagini visivamente simili (google, tineye, yandex, etc) Metadati Struttura dell'intestazione: esegue il dump della struttura dell'intestazione del file e visualizza una vista interattiva EXIF Full Dump: scansiona i metadati del file e raccoglie tutte le informazioni disponibili. Analisi miniature: estrae le miniature incorporate opzionali e le confronta con l'originale. Dati di geolocalizzazione: recupera i dati di geolocalizzazione opzionali e li mostra su una mappa del mondo

Ispezione Lente d'ingrandimento migliorata: lente d'ingrandimento con miglioramenti per una migliore identificazione dei falsi Istogramma dei canali: visualizzazione di singoli canali di colore o istogramma interattivo composito RGB Regolazioni globali: applica le regolazioni standard dell'immagine (luminosità, tonalità, saturazione, ...) Confronto di riferimento: apre una doppia vista sincronizzata per il confronto con un'altra immagine Dettaglio Gradiente di luminanza: analizza le variazioni di luminosità orizzontali/verticali dell'immagine Filtro bordi eco: utilizza i filtri derivati per rivelare le regioni artificiali fuori fuoco Soglia Wavelet: ricostruisce l'immagine con diverse soglie di coefficienti wavelet Divisione in frequenza: suddivide la luminanza dell'immagine in componenti ad alta e bassa frequenza. Colori Trame RGB/HSV: visualizzazione di trame interattive 2D e 3D dei valori dei pixel RGB e HSV Conversione spaziale: conversione dei canali RGB in spazi HSV/YCbCr/Lab/Luv/CMYK/Grigio Proiezione PCA: utilizza la PCA del colore per proiettare i pixel sulle componenti più salienti. Statistiche dei pixel: calcolo dei valori RGB minimi/massimi/medi per ogni pixel

Rumore Separazione del rumore: stima ed estrazione di diversi tipi di componenti del rumore dell'immagine Deviazione Min/Max: evidenzia i pixel che si discostano dalle statistiche minime/massime basate sui blocchi Valori dei piani di bit: mostra i singoli piani di bit per trovare modelli di rumore incoerenti Identificazione PRNU: sfrutta il rumore del modello del sensore introdotto da diverse fotocamere

JPEG Stima della qualità: estrazione delle tabelle di quantizzazione e stima della qualità JPEG salvata per ultima Analisi del livello di errore: mostra la differenza a livello di pixel rispetto a livelli di compressione

fissi Compressione multipla: utilizza un modello di apprendimento automatico per rilevare la compressione multipla Mappe fantasma JPEG: evidenziano le tracce dei diversi livelli di compressione nelle immagini di differenza Manomissione Miglioramento del contrasto: analizzare la distribuzione del colore per rilevare i miglioramenti del contrasto Falsificazione di copia e spostamento: utilizzo di descrittori di caratteristiche invarianti per il rilevamento di aree clonate Splicing composito: sfruttare le statistiche DCT per il rilevamento automatico delle zone di splicing Ricampionamento dell'immagine: stima dell'interpolazione dei pixel 2D per rilevare le tracce di ricampionamento Varie Filtraggio mediano: rileva le tracce di elaborazione lasciate dal filtraggio mediano non lineare Mappa degli illuminanti: stima della direzione della luce locale della scena sulle superfici 3D stimate Pixel morti/caldi: rileva e corregge i pixel morti/caldi causati da imperfezioni del sensore Decodificatore stereogrammi: decodifica delle immagini 3D nascoste negli autostereogrammi

Snap-Scraper Snap Scraper è uno strumento di intelligence open source che consente agli utenti di scaricare i media caricati sulla Snap Map di Snapchat utilizzando una serie di coordinate di latitudine e longitudine. Il programma può al momento essere usato solamente su MacOs avviando il file binario contenuto nella pagina di GitHub. Avviare il file binario aprirà una finestra del terminale avviando il programma. Selezionando la funzione principale verrà chiesto all'utente di inserire latitudine e longitudine del luogo in relazione al quale estrarre i media da Snapchat, con la possibilità di scegliere l'ampiezza del perimetro dell'area di ricerca. (Il readme consiglia di copiare latitudine e longitudine direttamente dall'url di map.snapchat.com dopo una ricerca del luogo, il che vanifica quasi del tutto l'utilità del tool in questione). Viene presentata poi a schermo una lista dei media e informazioni relative a esso, tra cui:

- URL del media
- Orario
- Durata dello snap
- ID dello snap
- Tipo di media dello snap
- Luogo

Spiderfoot Spiderfoot è uno strumento di automazione per la ricerca OSINT che si pone l'obiettivo di raccogliere nel web quante più informazioni possibili relative ad un utente, cercando di estrarre valore informativo da ogni piccola traccia. Si integra con quasi tutte le fonti di dati disponibili e utilizza una serie di metodi per l'analisi dei dati, rendendoli facili da navigare. SpiderFoot è dotato di un server web integrato per fornire un'interfaccia web-based, ma può anche essere utilizzato completamente tramite la riga di comando. È scritto in Python 3 e ha licenza MIT. Con il comando sudo spiderfoot -l 127.0.0.1:80 si avvia l'interfaccia web accessibile tramite browser in localhost. Qui si può dare il nome a una nuova scansione e indicarne il target. A seconda dell'input fornitogli seleziona automaticamente i moduli da attivare, in modo ottimizzare e rendere più efficienti le ricerche. In alternativa, i moduli possono essere abilitati dall'utente a seconda delle proprie necessità. Si può scegliere il livello di ricerca tra quattro tipologie di scansioni:

- Passive: vengono raccolte quante più informazioni possibili evitando rapporti diretti con il sito o gli account in possesso dal target, in modo tale da proteggere la propria identità investigativa.
- Investigate: vengono effettuate una serie di scansioni circa la vulnerabilità e la pericolosità del target.
- Footprint: si identifica la topologia di rete del target e vengono raccolte generiche informazioni a partire dal web e dai motori di ricerca, utili ad eseguire basiche operazioni di indagine.

- All: consigliabile per quando si necessita di informazioni dettagliate, vengono consultate tutte le possibili risorse e fonti d'informazione disponibili, direttamente o indirettamente collegabili al target. Per questo motivo, i tempi di elaborazione possono essere lunghi.

Una volta specificato il target di ricerca insieme a tutti i relativi parametri, Spiderfoot inizia la sua indagine raccogliendo quante più informazioni possibili. La natura di queste informazioni è molto disparata: indirizzi IP, nomi di dominio, indirizzi e-mail, numeri di telefono, nomi reali, nomi host, sottoreti di rete, ASN e altro. Al termine della ricerca, i risultati vengono mostrati tramite un semplice elenco oppure ordinatamente rappresentati in un grafico di nodi, insieme a tutte le entità, i collegamenti e le relazioni trovate tra di essi. È anche possibile esportare i dati in formato Excel o csv. Tutte le potenzialità, gli impieghi e le principali funzionalità di questo strumento possono essere riassunte nella seguente lista:

- Elaborazione dei dati: l'obiettivo primario di questo strumento è quello di estrarre quante più informazioni possibili da ciascuna scansione. Per questo motivo, ogni dato raccolto viene elaborato da più moduli, in modo da estrarre quanto più valore possibile.
- Semplicità e Velocità: grazie all'utilizzo di una semplice interfaccia utente accessibile da qualsiasi browser web, Spiderfoot risulta intuibile e di facile utilizzo. L'utilizzo di questa interfaccia permette di velocizzare sia l'avvio delle scansioni che la navigazione tra i risultati delle ricerche.
- Dark web: l'integrazione con la rete TOR permette di scansionare in maniera anonima e sicura anche i siti contenuti nel dark web.
- API: come alternativa di utilizzo all'interfaccia web, Spiderfoot rilascia una serie di API utili per eseguire scansioni, interrogare dati e svolgere altre operazioni direttamente nel codice di un qualsiasi progetto.

TheHarvester The Harvester è un strumento OSINT per la ricognizione che permette di recuperare informazioni su un dominio target utilizzando diverse fonti di informazioni per raccogliere risultati e determinare il perimetro di un'azienda. The Harvester raccoglie e-mail, sottodomini, IP e URL. È preinstallato in Kali Linux. Alcuni moduli del programma richiedono delle chiave API per avviare una ricerca, tra cui:

- bingapi
- hunter
- intex
- securityTrails
- shodan

TinEye TinEye è un motore di ricerca inversa di immagini gratuito per uso non commerciale. È possibile caricare un'immagine sul sito per scoprire da dove proviene, come viene utilizzata, se esistono versioni modificate dell'immagine o per trovare versioni a più alta risoluzione. TinEye utilizza una tecnologia di riconoscimento delle immagini invece che basarsi su parole chiave o metadati. Quando si invia un'immagine da ricercare, TinEye crea una firma digitale unica e compatta o "impronta digitale" per l'immagine utilizzando il riconoscimento delle immagini, quindi confronta questa impronta digitale con ogni altra immagine nell'indice per trovare delle corrispondenze. TinEye non trova in genere immagini simili (cioè un'immagine diversa con lo stesso soggetto), ma trova corrispondenze esatte, comprese quelle che sono state ritagliate, modificate o ridimensionate.

Gli utilizzi di TinEye sono molteplici, tra cui:

- Scoprire la provenienza di un'immagine o ottenere maggiori informazioni su di essa.
- Rintracciare la prima comparsa di un'immagine online
- Trovare versioni ad alta risoluzione di un'immagine
- Individuare le pagine web che fanno uso di un'immagine
- Scoprire versioni modificate o editate di un'immagine
- Aiuto per l'attribuzione delle immagini

Esiste una versione a pagamento di TinEye per uso commerciale, che consente di acquistare le ricerche di TinEye. La versione commerciale di TinEye include un'interfaccia utente per facilitare le ricerche e un'API per integrare TinEye con il proprio sito web o sistema.

Twayback Scarica i tweet cancellati dagli utenti salvati nella Wayback Machine. Può scaricare alcuni o tutti i Tweet cancellati di un utente. Consente di estrarre il testo dei tweet in un file di testo. Ha la possibilità di fare lo screenshot dei Tweet cancellati usando Playwright. (Gli screenshot per i tweet precedenti al 2016 non funzionano).

Consente di specificare un intervallo di tempo personalizzato per restringere la ricerca dei Tweet eliminati archiviati tra due date. (L'intervallo di date personalizzato non riguarda la data di creazione dei Tweet, ma piuttosto la data di archiviazione). Distingue tra account attivi, sospesi o che non esistono più. (se un account viene sospeso o non esiste più, tutti i suoi Tweet sono considerati eliminati). Consente di sapere se i tweet archiviati di un account di destinazione sono stati esclusi dalla Wayback Machine. Salva un registro degli URL dei tweet eliminati, nel caso in cui si desideri visualizzarli sulla Wayback Machine.

La qualità dei file HTML dipende dal modo in cui la Wayback Machine li ha salvati. Alcuni sono migliori di altri. Lo strumento è ottimizzato per tweet che non contengono video o immagini, in fase di prova sono infatti stati riscontrati problemi con la parte del programma responsabile degli screenshot.

Toutatis Toutatis è uno strumento che consente di estrarre informazioni dagli account di Instagram, come e-mail, numeri di telefono e altro ancora.

Per far funzionare il tool è necessario copiare il cookie session id dal browser mentre si naviga sul sito di Instagram mentre si è loggati con un account, le informazioni estratte possono essere:

- Nome dell'utente
- userID
- Se l'account target è verificato
- Se l'account target è aziendale
- Se l'account target è in stato privato
- Numero di account che seguono il target
- Numero di account seguiti dal target
- Numero di post del target
- Numero dei tag nei post del target
- url esterno
- Numero di post IGTV
- Biografia del target
- Email pubblica
- Telefono pubblico
- Email dell'account offuscata
- Telefono associato all'account offuscato
- url della foto profilo

L'utilità del tool è limitata, poiché gli utenti di instagram possono scegliere di rendere private molte delle informazioni che possono risultare utili. In alcuni casi l'utilizzo del tool fornisce le stesse informazioni che può dare il normale utilizzo dell'applicazione.

4.3.9 Descrizione del processo di analisi sui social network

Per provare l'efficacia di questi strumenti in ambito di analisi sui social network, nel contesto di questo Group Project è stato richiesto che venissero raccolte quante più informazioni possibili avendo come unico punto di partenza un indirizzo email, in questo caso quello universitario del professor Fausto Marcantoni (fausto.marcantoni@unicam.it). Tutti i tools sono stati utilizzati in un ambiente virtuale Kali Linux.

Per prima cosa si è voluto verificare l'appartenenza del dominio "unicam.it"

```
$ host unicam.it
unicam.it mail is handled by 5 ALT2.ASPMX.L.GOOGLE.COM.
unicam.it mail is handled by 5 ALT1.ASPMX.L.GOOGLE.COM.
unicam.it mail is handled by 1 ASPMX.L.GOOGLE.COM.
unicam.it mail is handled by 10 ASPMX3.GOOGLEMAIL.COM.
unicam.it mail is handled by 10 ASPMX2.GOOGLEMAIL.COM.
```

Poiché il dominio è stato creato partendo da Google Domains, l'indirizzo email si può considerare l'indirizzo come un indirizzo Google.

Alla luce di ciò, l'indirizzo è stato quindi usato come input in GHunt, da questo tool è stata ricavata attività del target su Google Maps.

Maggior parte delle informazioni trovate partendo dall'indirizzo mail sono però state trovate utilizzando il tool Spiderfoot. Esso infatti è stato capace di trovare una serie di account legati alla mail tra cui account: Pinterest, Instagram, Slideshare, Gravatar eVivino. Un'ottima caratteristica di Spiderfoot è quella di dare la possibilità all'utente di segnalare alcuni risultati come falsi positivi, in questo caso un account Pikebike appartentente a un utente geolocalizzato a Brooklin è stato scartato dal conto.

L'immagine di profilo usata sia su Pinterest, che Instagram, che Vivino è stata messa in input al servizio online PimEye, che ha permesso di trovare il sito del professore con il solo utilizzo dei tools.

Spiderfoot ha inoltre sottoposto l'indirizzo email al servizio Have I Been Pwned dove ha trovato il suddetto tra dati compromessi da violazioni dei dati, tra cui un account di Adobe, PeopleDataLabs e verifications.io. Utilizzando alcuni servizi a pagamento come Intelligence X, si potrebbero ricercare le password utilizzate in questi account.

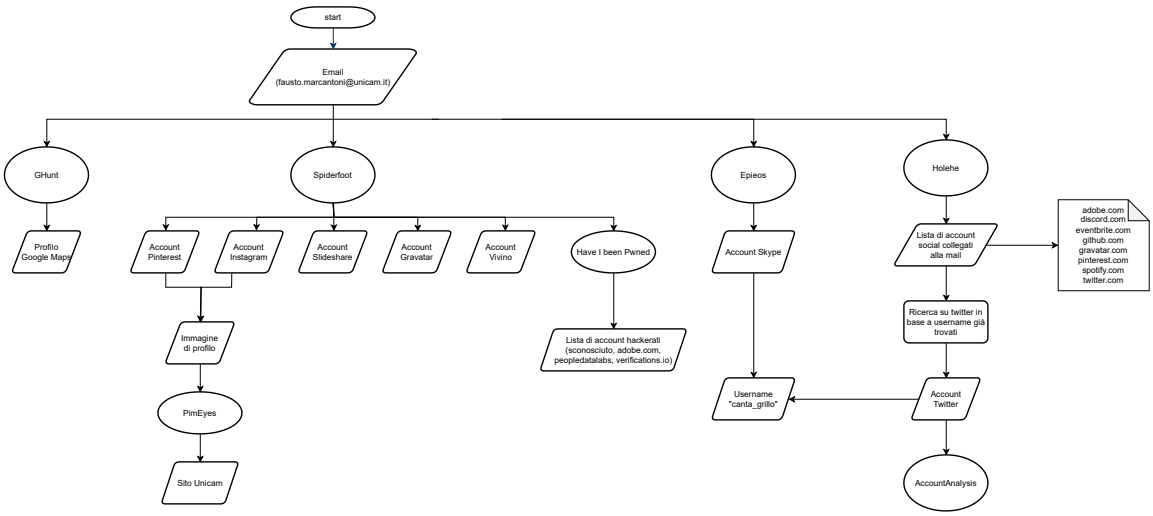
Mettendo invece in input l'indirizzo email in Holehe è stato è stata trovata una lista di account social collegati alla mail tra cui Adobe, Discord, EventBrite, GitHub, Gravatar, Pinterest, Spotify, Twitter.

Per ricercare gli account associate alla mail, Holehe utilizza un tecniche come di consulto di registri pubblici e la funzione di recupero password di alcuni siti.

Poiché non è possibile trovare un account Twitter in base a un indirizzo mail è stata fatta una ricerca manuale sul sito per trovare il account. Qui è stato trovato un account sicuramente appartenente al target, poi sottoposto ad analisi con accountanalysis. L'account Twitter presenta un handle con un username particolare, immaginando che possa essere stato utilizzato in altri account è stato utilizzato come input per il tool Sherlock la ricerca non ha dato però risultati, ma l'utilizzo del sito epieos.com ha permesso inoltre di scoprire che all'indirizzo email è associato un account Skype con lo stesso handle.

Nonostante il target di questa analisi non rientri nella fascia demografica tipica dei social network, è stato possibile accumulare una buona quantità di informazioni utilizzando solamente strumenti che permettono il recupero automatico di informazioni.

Figura 6: Visualizzazione del processo



4.4 Analisi Lessicale

Uno degli usi più potenti nell'era dei social media è quello di aggregare grandi parti di testo, da qualsiasi provenienza di fonti,cultura e nazionalità. A livello **elementare** l'analisi lessicale può essere usata per mostrare i termini più utilizzati su Google, mentre a livello **superiore** può essere utilizzata per analizzare il significato dietro il linguaggio e dedurre informazioni riguardo le persone che utilizzano i social. Le informazioni si suddividono in[82]:

- caratteristiche demografiche
- classe sociale
- situazione economica
- livello di istruzione

Oltre alle capacità analitiche, i metodi analitici lessicali avanzati dipendono spesso dalla disponibilità di un corpus di base di riferimento. Per corpus, in questo contesto, non si intende semplicemente una grande raccolta di testi, ma un insieme completo di testi che fornisce la base per l'analisi descrittiva di una lingua[83].

4.4.1 Analisi delle parole chiavi

L'analisi delle parole chiave (in inglese *"Keyness analysis"*) è una misura della frequenza con cui una parola ricorre in una frase o testo.

4.4.2 Profilazione della frequenza

La profilazione della frequenza (in inglese *"Frequency profiling"*) è la capacità di distinguere un corpo di testo da un altro in base all'occorrenza delle parole chiave nei due contenuti[84], come una sorta di comparazione.

4.4.3 Cluster

Cluster è una sequenza di due o più parole che potrebbero non essere importanti ma che possono servire per il metodo Analisi delle parole chiavi[85]

4.4.4 Collocazione linguistica

La collocazione linguistica (in inglese *"collocation"*) indica la probabilità che due parole identificate nella Analisi delle parole chiavi si verifichino frequentemente insieme. La parola identificata si chiama "**nodo**"[85].

4.4.5 Analisi del sentiment

Serve per identificare termini o entità su cui una persona ha una generale opinione che non è condivisa da classi differenti (Es. Politica). Ha anche una funzione critica utile a classificare un'espressione come positiva, negativa o neutra, ma un eccessivo affidamento rischia di sopravvalutare il ruolo dei social media[86].

4.4.6 Analisi della posizione linguistica

L'analisi della posizione linguistica (in inglese *"Stance analysis"*) utilizza le preferenze linguistiche per modificare i valori sottostanti di un individuo o un'espressione di atteggiamento verso un determinato concetto.

4.4.7 Processo del linguaggio naturale

Le generazioni precedenti di ricercatori e analisti di intelligence dovevano affidarsi a traduttori e interpreti umani per elaborare grandi quantità di testo in altre lingue. I progressi tecnologici nell'analisi del testo e nell'elaborazione del linguaggio naturale hanno ridotto notevolmente questo onere e oggi è disponibile una serie di risorse per una traduzione e un'elaborazione più rapide di materiali in lingua straniera. Alcune risorse, come Google Translate, sono gratuite e open source e invitano gli utenti a proporre traduzioni migliorate per i testi generati dalle macchine, che a loro volta migliorano e perfezionano gli algoritmi nel tempo.

4.4.8 Machine learning

Il Machine learning è il processo, su cui tutti i metodi precedenti si affidano, di insegnare a un programma software a prendere decisioni indipendenti dall'uomo dopo che il processo decisionale desiderato è stato prima processo decisionale desiderato è stato prima modellato in modo estensivo per il programma.

Il Machine learning richiede che gli esperti di apprendimento automatico e di linguistica computazionale progettino inizialmente i parametri e "insegnino" adeguatamente al computer come riconoscere modelli linguisticamente rilevanti in un testo scritto[87].

4.5 Analisi Geospaziale

Tabella degli strumenti testati sull'analisi geospaziale

L'analisi geospaziale lavora spesso in combinazione con altri metodi per produrre un'immagine più ricca delle dinamiche sociali, militari e politiche rilevanti per la HUMINT.

Si è espansa in modo significativo anche con la creazione di nuove piattaforme di social media che possono collegare automaticamente un post o un tweet a un luogo specifico attraverso il cosiddetto "**geotagging**".

4.5.1 Geo-tagging

I "geo-tag" sono dati incorporati che contrassegnano la longitudine e latitudine di un determinato post. Nel geo-tagging viene utilizzato il **GPS**.

4.5.2 Geo-locating

Utilizzando programmi open-source come Google Earth e Google Maps, gli analisti possono localizzare punti di riferimento specifici.[88]

4.5.3 Geo-inference

La inferenza geospaziale (in inglese *"geo-inference"*) consente di geolocalizzare un utente senza informazioni esplicitamente "geo-taggate". Alcuni siti web registrano la posizione dell'utente per personalizzarne l'esperienza, il che lascia "contenuti sensibili alla posizione" nella cache del browser dell'utente, infatti le parti interessate possono accedere alle geo-localizzazioni lasciate nella cache utilizzando "canali secondari" per determinare la posizione dell'utente, in particolare su piattaforme come Twitter.

4.5.4 Geo-referencing

La referenziazione geospaziale (in inglese *"geo-referencing"*) associa un oggetto a posizioni nello spazio fisico, ha principalmente scopi nell'ambito di ricerca militari e d'intelligence e serve ai sistemi di informazione per associare una mappa fisica o immagini a una mappa con posizioni spaziali[89].

4.5.5 Applicare strumenti di analisi geospaziale

Uno degli usi più potenti è quello di contestualizzare le informazioni su un determinato fattore e i suoi effetti sulle infrastrutture e sulla popolazione in tempo reale.

Alcuni software consentono agli utenti di scaricare mappe disponibili pubblicamente e supportate dalla piattaforma Google e di sovrapporre specifici fattori di interesse a fattori come la composizione etnica e religiosa di una regione per mostrare le interazioni tra geografia e religiosità in un'area. L'analisi geospaziale, dei social network e lessicale, abbinata a un elenco sempre più ampio di piattaforme software pubbliche e private, aiuta sempre di più a a dare un senso ai problemi dei big data. Le possibilità di apprendimento automatico e di elaborazione del linguaggio naturale sono enormi per la raccolta di OSINT.

Allo stesso tempo, gli analisti dell'intelligence continueranno a svolgere un ruolo fondamentale nel determinare come collegare le informazioni offerte da questi metodi in maniera in modo convincente e affidabile.

4.6 Elaborazione di dati geospaziali e utilizzo di strumenti

La definizione **GEOINT** sta per "GEOspatial INTelligence", utilizzando le prime tre lettere di ogni parola ("geo" e "int") per creare l'acronimo GEOINT ed indica una disciplina che comprende lo sfruttamento e l'analisi delle immagini e delle informazioni geospaziali per descrivere, valutare e rappresentare visivamente le caratteristiche fisiche e le attività geograficamente riferite sulla Terra.

Il termine GEOINT[90] è stato creato da un'agenzia di governo negli USA per scopi militari, infatti il termine è stato coniato da un tenete dell'aereo nautica James Clapper che definì la GEO INT in questa nota:

"La GEOINT comprende tutti gli aspetti delle immagini e delle informazioni e servizi geospaziali. Include, ma non si limita all'analisi di immagini letterali, dati geospaziali e informazioni tecnicamente derivate dall'elaborazione, dallo sfruttamento, dall'analisi letterale e non letterale di prodotti fusi spettrali, spaziali e

temporali. Questi tipi di dati possono essere raccolti su bersagli stazionari e in movimento tramite strumenti elettro-ottici, radar ad apertura sintetica (SAR), programmi di sensori correlati e mezzi non tecnici (comprese le informazioni geospaziali acquisite dal personale sul campo)"

La GEOINT è cresciuta al di là della sua iniziale focalizzazione negli Stati Uniti, tanto da essere che l'analista GEOINT è un titolo di lavoro dell'esercito degli Stati Uniti e sta diventando uno standard accettato a livello globale.

La GEOINT combina diverse discipline come la mappatura, la cartografia, l'analisi delle immagini e l'intelligence delle immagini. Sebbene sia normalmente associata a un contesto militare, il fatto è che sempre più organizzazioni del settore civile e privato che operano in settori come le telecomunicazioni, i trasporti, la salute e la sicurezza pubblica e il settore immobiliare utilizzano l'intelligence geo-spaziale per migliorare la qualità della vita quotidiana.[91]

Il principio di base della GEOINT consiste nell'organizzare e combinare tutti i dati disponibili in base alla loro posizione geografica sulla Terra, per poi sfruttarli al fine di preparare prodotti facilmente utilizzabili da degli utilizzatori.

Nel corso degli anni si è creata una certa confusione tra le competenze della disciplina **IMINT** e della **GEOINT**. Capita di confondere questi 2 tipi di intelligence ma è considerato un errore concettuale, perché la GEOINT è un modo di analizzare e correlare i dati e le informazioni che provengono dalle discipline di raccolta, essa presenta i dati intelligence di varie fonti, legandoli ad una posizione sulla terra in un tempo specifico; quindi, la GEOINT può essere definita come un prodotto intelligence. Mentre la IMINT, acronimo di IMagery INTelligence, è l'attività di raccolta di informazioni mediante l'analisi di fotografie aeree o satellitari.[92]

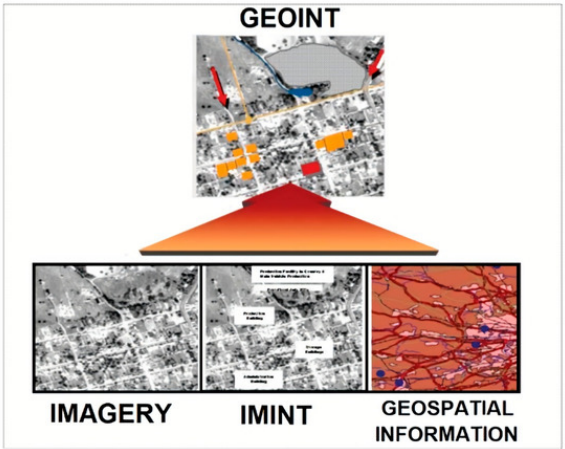


Figura 7: GEOINT integration

L'**analista di immagini dell'intelligence geo-spaziale** è responsabile dell'analisi di immagini fornendo al personale dell'esercito informazioni critiche sulle forze nemiche, un analista GEOINT può applicare le stesse competenze alle forze dell'ordine, alle agenzie umanitarie globali, ai sistemi di risposta alle emergenze, alle società di consulenza investigativa e di analisi forense e alle società di consulenza sulla sicurezza.

Alcuni dei settori beneficianti della GEOINT sono:

- **Settore commerciale** - La GEOINT sta dando forma a informazioni censuarie, storiche, meteorologiche e geologiche per molti usi commerciali. I settori dei viaggi, dell'immobiliare e del petrolio e del gas beneficeranno di dati GEOINT accurati .

- **Settore militare** - La GEOINT aiuta a pianificare ed eseguire le attività, a stabilire le reti di distribuzione e le operazioni RSOI (ricezione, stazionamento, movimento successivo e integrazione).
- **Settore automobilistico** - I veicoli autonomi dipenderanno da dati GEOINT altamente precisi per la sicurezza.
- **Settore della sicurezza** - I dati GEOINT aiuteranno le forze dell'ordine a combattere gli incendi in diverse aree geografiche come boschi e a mantenere i vigili del fuoco al sicuro grazie a dati precisi sulla posizione e sulle condizioni meteorologiche, anche nell'identificazione di aree geografiche dove sono accaduti o staranno per accadere eventi o attività criminali, cercando di svolgere un'attività preventiva.
- **Settore agricolo** - I dati GEOINT possono massimizzare la salute delle piante e ridurre al minimo l'uso di fertilizzanti su ampie superfici grazie al loro preciso livello di dettaglio localizzativo.

4.7 Descrizione del processo GEOINT

Il processo di geo-localizzazione si sviluppa dopo che l'analista di immagini ha elaborato e spacchettato la fonte iniziale trovando tutte le informazioni da cui l'analista geo-spaziale riuscirà a processare selezionando i diversi tools di lavoro più adatti al contesto.

Il processo di geo-localizzazione si divide in diversi step:

- Image analysis
- Reverse image
- Geolocation (coordinate lat, long)
- Geo mapping

Questi 4 step verranno sviluppati con l'ausilio di appositi tools d'elaborazione.

1. Nelle prime fasi del processo si dovrà, una volta ottenuto il file (immagine o video) su cui lavorare, si dovrà **verificare** se il file non sia stato corrotto e provenga da una fonte attendibile, quindi verificare l'immagine attraverso l'URL con tools come UrlScan che permette di ricavare meta dati e certificati da un determinato url, questi dati agevoleranno il compito dell'analista OSINT nella validazione dell'immagine su cui si effettuare la ricerca, una volta validata l'immagine si otterranno dei meta dati più specifici attraverso tools come Meta2Go, pic2Map ottenendo così Geo-tags(latitudine e longitudine) se la foto ne ha, anche timestamp del momento in cui è stata scattata l'immagine e in alcuni casi servirà sapere anche il modello della camera con cui è stata scattata. Se l'immagine è stata presa dal browser sarà più difficile ottenere i meta dati, almeno che l'utente che ha effettuato l'upload non li abbia inseriti manualmente

   Si possono trovare i meta dati anche di file video presenti su YouTube.

2. In seguito dovrà si dovrà effettuare **reverse image**, tecnologia dei motori di ricerca che prende un file di immagine come query di input e restituisce i risultati relativi all'immagine, in modo manuale facendo l'upload delle immagini su dei tools come tinyEye o Google Lens, oppure alcuni browser forniscono delle estensioni come Invid verification tool che permette con un click di verificare l'immagini su diversi motori di ricerca come Yandex, Bing, Google e altri. Per i contenuti presenti su YouTube si possono utilizzare tools come Youtube GeoFind oppure YT DataViewer. Nell'analisi di un video la divisione in frame agevoleranno il processo.

   Il processo di reverse image fornisce inoltre altri funzionalità:

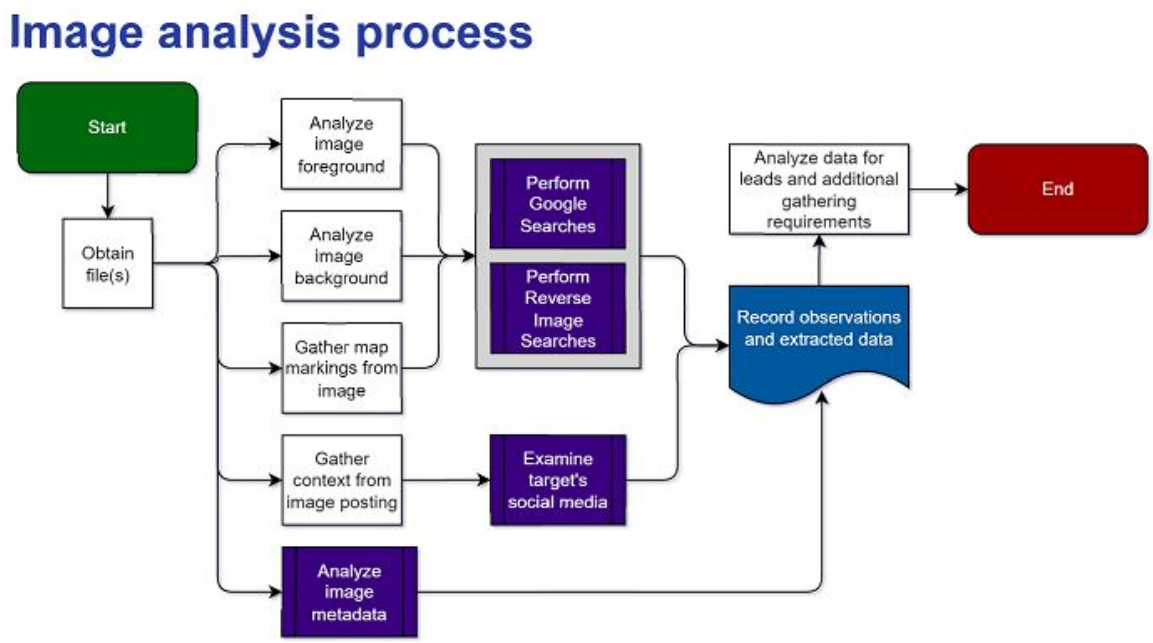


Figura 8: IMINT process

- Individuare le informazioni sulla fonte di un'immagine.
- Trovare versioni ad alta risoluzione delle immagini.
- Trovare informazioni su prodotti e altri oggetti non identificati.

Una "buona norma" è quella di ritagliare la foto e scegliere solo l'area selezionata ed esaminarla.

3. In seguito si **analizzerà la foto**, dopo aver selezionato la versione con la qualità migliore, questo step consiste nell'analisi di tutti i dettagli sia nello sfondo che in primo piano:

   - Strutture
   - Condizioni meteorologiche ed orario
   - Oggetti atti alla segnaletica
   - Paesaggio

   Per visionare al meglio i dettagli nel paesaggio si possono utilizzare diversi tools come Map channel per un visione completa con ogni versione(satellitare, mappa, street view) del luogo, Wikimapia, Google Earth che fornisce immagini 3D e immagini storiche del luogo (no ultima versione), Bing Maps che fornisce maggiori aggiornamenti nelle immagini satellitari e il famoso motore di ricerca russo Yandex che fornisce un servizio di mapping.

   Durante l'analisi del paesaggio bisogna far attenzione se l'immagine lo permette al meteo e al vento. Per visionare dati passati del meteo, il tool Wolframalpha fornisce questa funzione di recupero dei meta dati meteorologici attraverso una query utilizzando il linguaggio naturale, per il vento invece il tool ads-B Exchange fornisce il movimento dei venti, questo caso di studio si potrà analizzare solo nel caso che nell'immagine mostri qualche oggetto che si muova in una determinata direzione a causa del vento.

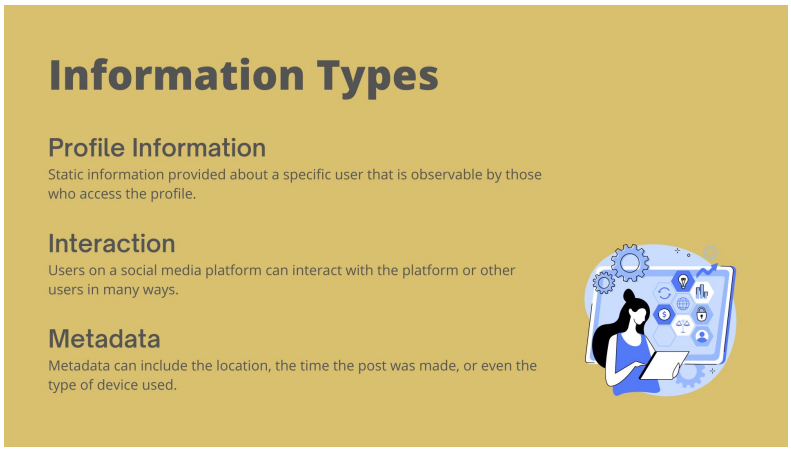


Figura 9: Tipi di informazioni

4. Infine, come ultimo step si potranno **creare delle mappe** con i dati geo localizzati, inserendo i geo tags e altri dati ottenuti in file csv o json che processati da tools come MapMySheet o Batch2Geo creeranno una mappa e mapperanno le posizioni desiderate.

4.7.1 Strumenti per l'analisi geospaziale

Telegram nearby map Telegram nearby map è un strumento OSINT basato su nodeJS che utilizza OpenStreetMap e la libreria ufficiale(API ufficiali) di telegram per trovare la posizione degli utenti nelle vicinanze.

Youtube-geofind Youtube-geofind è un strumento presente nel web per la ricerca su YouTube di video etichettati geograficamente per canale, argomento e località.

What is where What is where è un potente strumento di ricerca basato su mappa, di tipo POI search (point of interest).

Yandex Yandex images è uno strumento di ricerca di foto, anche tramite foto, adatto per reverse image. Può essere utilizzato anche come estensione web o come web app.

Live ua map Live Universal Awareness Map ("Liveuamap") è uno strumento di geolocalizzazione e mappatura nel settore delle notizie e delle informazioni globali, dedicato alla segnalazione di una serie di eventi importanti, tra cui conflitti, questioni relative ai diritti umani, proteste, terrorismo, dispiegamento di armi, questioni sanitarie, disastri naturali e storie meteorologiche, tra le altre cose, provenienti da una vasta gamma di fonti.

Echosec Echosec è una piattaforma di individuazione di dati basata sul web, che consente alle organizzazioni di rilevare i dati online per ottenere informazioni sulle minacce. Aggregando e mappando contenuti provenienti da centinaia di fonti, tra cui social media, blog e notizie, Echosec analizza e monitora i dati open source in tempo reale per la protezione dei marchi, il monitoraggio degli eventi, la salvaguardia dei dirigenti, la vendita

al dettaglio e la finanza. Echosec utilizza il machine learning per filtrare miliardi di post con il rilevamento delle parole chiave e il riconoscimento delle immagini, fornendo avvisi sulla priorità al team.

SunCalc SunCalc è uno strumento che mostra il movimento del sole e le fasi della luce solare durante il giorno in una determinata località. È possibile vedere le posizioni del sole all'alba, all'ora specificata e al tramonto.

GeoWifi Geowifi è uno strumento di ricerca di dati di geolocalizzazione WiFi per BSSID e SSID su diversi database pubblici. Vengono analizzati diversi database dopo aver inserito il BSSID (MAC) E il SSID e vengono restituiti i relativi geotag se presenti nei vari database.

Maphub MapHub è uno strumento di geolocalizzazione di eventi nel mondo, in particolare della guerra in Ucraina. In alcuni casi questi eventi segnalati nella mappa vengono accompagnati da dei post pubblicati nei vari social (facebook, twitter, instagram etc..)

Google dork Le Google Dorks o "Comandi di Google" costituiscono un metodo per migliorare i risultati della ricerca. Per trovare risultati specifici, è sufficiente inserire, all'interno della query, qualche "keyword" particolare o simbolo in più in modo da essere più selettivo.

Google lens-images Google immagini è uno strumento che permette di matchare un'immagine con altre, così fornendo informazioni e/o rintracciare l'immagine. Viene trascinata un'immagine nel box di ricerca, si cercano sul web le immagini visivamente simili a quella caricata.

Sherloq Sherloq è un programma che offre una raccolta di strumenti open source per l'analisi forense di immagini digitali. Una volta avviato il programma si mostra come completo di gui. Permette di caricare un'immagine da analizzare e scegliere i vari tipi di analisi alle quali sottoporre le immagini.

BatchGeo BatchGeo è uno strumento di geo mapping che permette di inserire in una mappa,fornita dal servizio di google Maps, dei POI (point of interest). Questi punti dovranno essere inseriti inizialmente in una Google Sheet oppure inseriti manualmente nel form fornito nella pagina iniziale, inserendo nome e coordinate geospaziali ed altre informazioni utili (esplicitate nel foglio di esempio) e poi verranno processate dal tool e geo mappate con dei marker nella mappa.

Pic2Map Pic2Map è uno strumento di visualizzazione di dati EXIF online con supporto GPS che consente di localizzare e visualizzare le foto su una mappa. Il sistema utilizza i dati EXIF, disponibili in quasi tutte le foto scattate con fotocamere digitali, smartphone e tablet. Anche senza dati GPS, Pic2Map funge comunque da semplice e completo visualizzatore online di dati "EXIF", acronimo di Exchangeable Image File, un formato standard per la memorizzazione di informazioni di interscambio nei file di immagini di fotografia digitale che utilizzano la compressione JPEG.

World webcams World webcams è uno strumento di geolocalizzazione che permette di visualizzare le immagini di centinaia di webcam, telecamere di tutto il mondo, compresa una descrizione della posizione della webcam.

MapChannels Map channels è uno strumento che permette la visualizzazione di mappe della posizione desiderata, il suo punto di forza è quello che permette la visione di più prospettive e modalità di visualizzazione nella stessa schermata.

Gpsvisualizer GPS Visualizer è uno strumento che permette di mappare dei luoghi. I suoi punti di forza sono la semplicità e la flessibilità in termini di input e l'enorme numero di opzioni per quanto riguarda l'output.

Ads-B Exchange ADSBexchange è uno strumento di geolocalizzazione e la più grande fonte pubblica al mondo di dati di volo non filtrati. L'accesso ai dati di tracciamento dei voli di tutto il mondo apre un mondo completamente nuovo di monitoraggio dei ricercatori OSINT.

GoogleMySheet GoogleMySheet è un strumento di geomapping che permette di inserire in una mappa, fornita dal servizio di Google Maps, dei POI (point of interest). Questi punti dovranno essere inseriti inizialmente in una Google Sheet, inserendo nome e coordinate geospaziali ed altri informazioni utili e poi verranno processate dal tool e geo mappate con dei marker nella mappa.

GpsJam Gps Jam è un strumento che fornisce una mappa mondiale delle interferenze nelle vari luoghi. Le interferenze possono essere più o meno forti, questo viene registrato attraverso un codice colori.

Carnet CarNET è uno strumento che permette di rilevare la marca, modello, generazione, colore e angolazione partendo da un'immagine di un auto. Le API di questo strumento sfruttano tecniche di *computer vision* e *deep learning*.

Rare earth element map La Rare Earth Element map è una mappa che identifica, con il supporto di Open Street View Map, tutte le aree geografiche che presentano gli elementi rari.

Wigle Wigle è un strumento di geo mapping che permette di visualizzare la posizione e le informazioni delle reti wireless di tutto il mondo contenute in un database centrale che è possibile mappare, interrogare e aggiornare.

Youtube Metadata YouTube MetaData è uno strumento di analisi dei video presenti su youtube, restituendo tutti i metadati presenti nell'url del video caricato.

Url scan urlscan.io è uno strumenti che effettua la scansione e analisi dei siti web. Quando un URL viene inviato a urlscan.io, un processo automatizzato naviga nell'URL come un normale utente e registra l'attività che la navigazione della pagina crea. Questo include i domini e gli IP contattati, le risorse (JavaScript, CSS, ecc.)

Wolfaramalpha Wolfram è uno strumento e "motore computazionale di conoscenza" che "decodifica ed elabora" la richiesta dell'utente, intrecciando i dati a sua disposizione (base di conoscenza), mostrando definizioni o eseguendo calcoli e confronti a seconda dei casi.

Foto forensic Foto forensic è uno strumento di analisi delle foto tramite url o upload diretto da file system. Permette di trovare: metadati, informazioni dettagliate, pixel nascosti, miglioramento colori icc+, stringa immagine, dettagli sul formato ELA (error level analysis)

Geolocation Estimation Geolocation Estimation è uno strumento che permette di trovare la posizione di un luogo inserendo l'immagine corrispondente. Analizzando i punti salienti della foto e inserendo manualmente dove si potrebbe trovare la foto cerca di localizzarla in modo più preciso possibile.

5 Privacy, sicurezza e legge

L'OSINT automatizzata può essere descritta come la raccolta di dati disponibili pubblicamente con l'ausilio di software specializzati o applicazioni web, è spesso considerata un tipo di sorveglianza non intrusiva, poiché i dati sono disponibili a chiunque utilizzando motori di ricerca come Google o acquistando dati.

Tuttavia, l'OSINT tradizionale si è evoluta in una pratica professionale e invasiva, non consiste più nel consultare gli elenchi telefonici o cercare dati su Internet utilizzando un motore di ricerca. Al contrario, è possibile utilizzare software dedicato che interroga centinaia di fonti contemporaneamente, tra cui i dati dei servizi di social media, i dati di localizzazione generati dalle pubblicità sulle app, sui telefoni cellulari e i dati trapelati dagli utenti. Gli individui condividono regolarmente informazioni personali online, in particolare nei social network, che vengono archiviate sotto forma di dati digitali in database o nel cloud. Questo, a sua volta, ha portato a nuove percezioni su come questi dati personali possano essere utilizzati per scopi di sicurezza e protezione.

Poiché Osint si appropria di informazioni pubbliche, gli utenti devono essere messi al corrente dei rischi che si corrono inserendo troppe informazioni o comunque utilizzare la rete in modo ingenuo. Osint oltre ad essere uno strumento per trovare vulnerabilità o informazioni sensibili, può essere utilizzata anche per evitare attacchi informatici. Le aziende utilizzatrici di questi tools Osint ottengono un profitto elaborando dati pubblici che verranno poi messi a disposizione degli utenti dietro il pagamento di un corrispettivo oppure in altri casi vengono messe a disposizione ad agenzie di sicurezza e di intelligence o alla polizia, che probabilmente le utilizzano a fini investigativi.[93]

Mentre le aziende governative, come le forze dell'ordine o militari, ottengono queste licenze gratuitamente, per l'utilizzo di questi tools e per l'accesso a questi record di dati. Negli USA è in corso da tempo un dibattito sui tools Osint e sull'elaborazione dei dati di localizzazione. L' IC utilizza i dati di localizzazione dei singoli utenti senza mandato e autorizzazione, alcuni legislatori sostengono che sia necessaria un "autorizzazione" prima di procedere con l'analisi dei dati geolocalizzanti.

In Europa invece non è avvenuto questo dibattito in quanto l'Osint viene considerato uno strumento di sorveglianza non intrusivo, quindi L'IC può quindi utilizzare i vari tools entro i limiti della legge e nei principi di protezione dei dati.[94]

5.1 Responsabilità delle fonti aperte

Le informazioni open source hanno aumentato la gamma di strumenti di sicurezza a disposizione dei funzionari di sicurezza e di intelligence o degli agenti di polizia. Tuttavia, gli effetti collaterali di questo nuovo metodo di raccolta di informazioni dovrebbero essere bilanciati da una forma di responsabilità sufficiente sia in teoria che in pratica. A titolo di esempio, nella maggior parte delle società occidentali esiste una legislazione rigorosa per le intercettazioni telefoniche o di Internet, ma per i siti o le applicazioni di social network questo è meno frequente. L'uso dell'OSINF per scopi di intelligence ha conseguenze reali. Dal punto di vista dei diritti umani, questi effetti collaterali dovrebbero essere bilanciati. Si vuole quindi esaminare le responsabilità dello Stato per l'uso dell'OSINF.

Questa forma di responsabilità si caratterizza per l'attenzione allo stato di diritto e al buon governo, nonché per l'inclusione della società civile o della gente comune.[95] Quirine Eijkman e Daan Weggemans in "Open source intelligence and privacy dilemmas" suggeriscono che per arrivare a una forma di buon governo in relazione al bilanciamento dell'uso dell'OSINT da parte dei funzionari di sicurezza, il capo di un'agenzia di sicurezza e di intelligence, o i responsabili politici, non solo dovrebbero annunciare pubblicamente lo scopo della raccolta, dell'elaborazione, dell'estrazione o della condivisione dell'OSINF, ma ne limitare anche l'uso a minacce predefinite come la sicurezza nazionale (ad esempio per lo spionaggio informatico, il terrorismo internazionale). Inoltre, i legislatori internazionali e/o nazionali dovrebbero stabilire quali sono i confini (lo stato di diritto) e come gli interessati possono chiedere un risarcimento (meccanismi di responsabilità

interni o esterni). Infine, la progettazione del software che consente l'OSINT e allo stesso tempo enfatizza la responsabilità (data protection by design) dovrebbe essere modificata di conseguenza. Ciò include le tecnologie di rafforzamento della privacy o della trasparenza.[93]

Oltre alla necessità di proteggere gli interessi della sicurezza nazionale, che può interferire la trasparenza delle operazioni, esistono altri dilemmi di responsabilità in relazione all'OSINF. Garantire la responsabilità è più complesso se le informazioni non sono raccolte dall'agenzia di sicurezza stessa, ma da altri enti pubblici o privati. Non è raro che le agenzie di intelligence e di sicurezza condividano informazioni a livello nazionale e internazionale. Si tratta di uno sviluppo recente perché "tradizionalmente esiste una distinzione tra la raccolta di informazioni di intelligence per scopi di sicurezza nazionale e la raccolta di prove per le indagini penali, in quanto hanno scopi diversi".[96] Le agenzie di sicurezza e di intelligence preferiscono mantenere riservate le loro fonti, mentre le agenzie di polizia devono condividere informazioni con il consiglio di difesa. Ciononostante, l'OSINT utilizzata dalle agenzie di sicurezza e di intelligence viene raccolta ed elaborata anche da altre agenzie statali e talvolta condivisa con partner internazionali e questo ha influito sulla responsabilità nella pratica. Si pensi, ad esempio, al fatto che gli analisti di sicurezza sappiano quale sia la fonte originaria di un'informazione contenente dati personali, per non parlare del fatto che le persone interessate possano mai avere l'opportunità di accedervi o di correggerla.

Allo stesso modo, OSINF e OSINT sono raccolti anche da aziende private. I Wikileaks Global Intelligence Files, ad esempio, riflettono il modo in cui la società Stratfor fornisce intelligence a enti pubblici e privati, tra cui la Defense Intelligence Agency statunitense.[97] Una delle 5,5 milioni di e-mail di Stratfor pubblicate rivela l'esistenza di un sistema software predittivo TrapWire della TrapWire Inc. che combina le immagini delle telecamere a circuito chiuso e il riconoscimento delle targhe (Automatic number-plate recognition) raccolte nel dominio pubblico di due città americane.[98] Un'altra multinazionale della sicurezza, Raytheon, ha sviluppato un programma antisommossa che analizza i siti di social network e, sulla base dei risultati, è in grado di rintracciare le persone nel luogo in cui si trovano.[99] Evidentemente, le entità private sono in grado di vendere o condividere il software o le informazioni open source con le agenzie di sicurezza e di intelligence o con la polizia, che probabilmente le utilizzano a fini investigativi.[100] Questi sviluppi avvengono mentre le questioni relative alla responsabilità per l'uso dei dati personali memorizzati sui siti di social network rimangono irrisolte. Come conclude Ben Hayes, "dobbiamo (quindi) sviluppare gli strumenti e le comunità necessarie per metterli sotto controllo democratico".[101]

5.2 Le investigazioni digitali nella legislazione italiana

L'OSINT dal punto di vista giuridico non è coperto da una nozione giuridica univoca. Intervenendo su fonti aperte, ossia su fonti non sottoposte ad un regime di riservatezza, non c'è bisogno della collaborazione del titolare o di terzi per il cui recupero delle informazioni. La connotazione "aperta" delle fonti consente all'investigatore di non oltrepassare i limiti della legalità nella ricerca di informazioni. Se per un verso non è automatico il sillogismo tra fonte aperta e fonte affidabile, in quanto le fonti chiuse di solito sono connotate da un maggiore livello di affidabilità derivante dalla responsabilità di chi le cura, non è neppure automatico che la fonte aperta sia di per sé inaffidabile, soprattutto qualora sia validata da un'attività di cross-checking (ricerca di metodi alternativi all'originale per trovare la stessa "risposta").

Sempre da un punto di vista giuridico, parlare di fonti aperte non significa che si possa agire senza rispettare alcun paletto; infatti, il contenuto delle informazioni ricercate rientra – di regola – nella definizione di "dato personale", e dunque l'utilizzo di tecniche di OSINT solleva dei problemi in merito alla protezione accordata dal GDPR. Non ci sono invece limitazioni in relazione ai dati personali già diffusi da altre fonti e liberamente disponibili in rete, ed i dati già diffusi dallo stesso interessato (che possono essere lecitamente trattati ai sensi dell'art. 9, comma 2, lett. e) del GDPR); per questa tipologia di fonti sarà necessario accertarsi della liceità della fonte e della base giuridica in base alla quale viene comunicata.

L'Osint può essere utilizzato sia per scoprire informazioni come attaccante ma anche per compiere delle investigazioni digitali in ambito difensivo. Queste investigazioni sono tutelate sia dal codice penale che quello civile:

5.2.1 Penale

In materia penale, la prova scientifica si forma frequentemente sulla scena del crimine. Questa situazione ha portato a un rovesciamento di prospettiva per il difensore tenuto a ripensare sempre più spesso alla propria strategia difensiva. Il codice di procedura penale, infatti, agli artt.391 bis, offre molteplici strumenti per le investigazioni difensive finalizzate alla ricerca di elementi di prova: consente l'accesso alla documentazione, l'audizione delle persone informate dei fatti, l'accesso alla documentazione della pubblica amministrazione. Si considera il disposto dell'art.191 c.p.p., ai sensi del quale le prove acquisite in violazione dei divieti stabiliti dalla legge non possono essere utilizzate.

5.2.2 Civile

Il processo civile, invece, non ha regole specifiche per l'acquisizione della prova digitale simili a quelle introdotte nel processo penale con la l. 48/2008, ma ai sensi dell'art. 46 del Reg. UE 2014/910 (EIDAS[5]) ad un documento elettronico non sono negati gli effetti giuridici e l'ammissibilità come prova in procedimenti giudiziali per il solo motivo della sua forma elettronica. Ai sensi dell'art. 20, comma 1 bis, C.A.D., il documento informatico soddisfa il requisito della forma scritta e ha l'efficacia prevista dall'art. 2702 c.c. quando vi è apposta una firma digitale, altro tipo di firma elettronica qualificata o una firma elettronica avanzata. In generale, la validità del documento informatico che soddisfa il requisito della forma scritta e il suo valore probatorio, sono liberamente valutabili in giudizio in relazione alle sue caratteristiche oggettive di qualità, sicurezza, integrità e immodificabilità. Inoltre, ai sensi dell'art. 2712 c.c., le riproduzioni informatiche formano piena prova dei fatti e delle cose rappresentate, se colui contro il quale sono prodotte non ne disconosce la conformità ai fatti o alle cose medesime (implicito richiamo all'art. 115 c.p.c. sull'onere di specifica contestazione). Gli strumenti a disposizione del difensore sono molteplici ed in primis provengono dal GDPR[6]: l'art. 15 del Reg. UE 2016/679, consente all'interessato da un trattamento di ottenere dal titolare del trattamento la conferma che sia o meno in corso una manipolazione di dati personali che lo riguardano ed ottenerne l'accesso e la copia (i provider ed i servizi di messaggistica istantanea più diffusi, consentono di scaricare tutti i dati personali detenuti). l'art. 20 del Reg. UE 2016/679, attribuisce all'interessato il diritto alla portabilità dei dati, ossia il diritto di ricevere in un formato strutturato di uso comune e leggibile da dispositivo automatico, i dati personali che lo riguardano. In secondo luogo, è possibile utilizzare tutte le tecniche di OSINT, a seconda delle esigenze di ricerca.

Per avviare delle indagini difensive telematiche il difensore dovrà eseguire degli step, indipendentemente che si agisca in ambito penale o in ambito civile. Come primo step bisogna effettuare un'analisi forense, in cui si individua, conserva, protegge, estrae la documentazione e impiega il dato in un trattamento al fine di essere valutato, in questo caso si parla di strumenti hardware come smartphone o pc. Poi vengono analizzati i dispositivi IoT che circondano il soggetto. Con l'utilizzo del GPS si possono analizzare i dati e trovare la posizione in un luogo in una determinata ora. Il traffico dettato dai dispositivi IoT è tutelato dall'art. 132 Codice Privacy consentendo al difensore della persona sottoposta a indagini di richiedere direttamente al fornitore dei dati relativi alle utenze intestate al proprio assistito.

Affinché le attività di OSINT siano conformi alla normativa posta a protezione dei dati personali, il difensore è tenuto a verificare la legittimità e la qualità della fonte, minimizzare i dati raccolti e cancellare i

[5]EIDAS = Electronic IDentification authentication and signature

[6]GDPR = Il Regolamento generale per la protezione dei dati personali 2016/679 (General Data Protection Regulation)

dati non necessari, evitare la comunicazione dei dati a terzi, proteggerli e garantire l'esercizio dei diritti ai titolari degli stessi.

5.3 La giurisprudenza ed il web come fonte di prova documentale

Da un profilo giuridico, la materia delle fonti aperte nel processo penale, nonostante sia ancora nuova ed in continua evoluzione, ha comunque trovato alcuni recenti orientamenti utili a tracciarne le linee guida:

- La Corte Suprema
- La sezione del Lavoro (sentenza n. 2912/2004)
- Il trattamento dell'assimilabilità di una prova documentale (art 234 c.p.p., materiale proveniente dal web e prodotto in giudizio, "Le informazioni tratte da una rete telematica sono per natura volatili e suscettibili di continua trasformazione e a prescindere dalla ritualità della produzione, va esclusa la qualità di documento in una copia su supporto cartaceo che non risulti essere stata raccolta con garanzie di rispondenza all'originale o di riferibilità a un ben individuato momento")

Buona regola vuole, nelle investigazioni giudiziarie e nelle analisi di intelligence soggette a validazione, che l'acquisizione delle tante ricostruzioni rinvenute sulla rete venga, prima di tutto, richiamata riportando fedelmente il "link" a quel documento, precisando anche la data di consultazione della pagina. Ad esso andrebbero, poi, aggiunte tutte le cautele effettuate al fine di ricontrare l'attendibilità della notizia e della fonte che l'ha pubblicata, per poi integrare la veridicità dei contenuti con altre pratiche di polizia giudiziaria tradizionali (dalla consultazione delle fonti degli archivi informatici del Dipartimento di P.S., all'acquisizione formale di documenti di riscontro e arrivando all'interrogatorio delle persone che avevano pubblicato quelle notizie o che ne avevano conoscenza per ragioni professionali/personali).

5.4 L'OSINT sotto analisi

L' "analisi delle fonti aperte" fa parte integrante dell'attività di intelligence e si avvale di diversi strumenti:

- Mezzi di comunicazione di massa (giornali, riviste, televisione, radio e siti web)
- Dati pubblici (rapporti dei governi, piani finanziari, dati demografici, dibattiti legislativi, conferenze stampa, discorsi, avvisi aeronautici e marittimi)
- File multimediali (video, audio, fotografie e mappe)
- Informazioni provenienti da database istituzionali o privati come ad esempio Internet che offre diverse fonti da cui acquisire informazioni (siti web, blog, social network, forum, canali IRC, reti P2P, TOR, ecc.)

L'OSINT dovrebbe utilizzare esclusivamente fonti aperte ed ottenere, dunque, informazioni "non classificate" e cioè informazioni disponibili al pubblico, anche se non necessariamente ad alta divulgazione o ad accesso gratuito. Il problema, tuttavia, è che tale specifica azione informativa, caratterizzata da tempestività, aderenza e continuità, rischia di entrare in conflitto con la tutela di fondamentali diritti della persona, dalla "riservatezza" sino alla cd. "autodeterminazione informativa". La soluzione circa la legittimità o meno dell'attività di indagine in argomento dipende dal tipo di informazione estrapolata in modo occulto e, in particolare, dall'aspettativa di riservatezza che il soggetto ripone su di essa. Partendo da questa osservazione, privilegiato dal giurista, è possibile distinguere i dati pubblici, o di pubblico dominio, dai dati riservati. Nell'ambito dei social network, dati pubblici possono essere definiti tutti quei contenuti digitali postati volontariamente dall'utente sul proprio profilo personale e destinati ad un qualsiasi utente esterno. Facebook, ad esempio, consente agli iscritti di creare una propria pagina nella quale è possibile pubblicare immagini, filmati ed altri contenuti multimediali (l'accesso a questi contenuti è regolato attraverso impostazioni sulla privacy prestabilite dall'utente). Tuttavia, le informazioni e le fotografie pubblicate sul profilo e non sotto il controllo di filtri di visibilità non possono considerarsi riservate e non godono, quindi, della tutela della loro eventuale divulgazione ad opera di terzi. In altri termini, nel momento in cui si pubblicano informazioni e foto sul proprio profilo personale, rendendole accessibili a tutte le persone, si accetta il rischio che le stesse possano essere portate a conoscenza anche di terzi non rientranti nell'ambito delle "amicizie" accettate dall'utente. Riservate, invece, sono quelle informazioni che l'utente ritiene di voler condividere esclusivamente con la sua cerchia di "amici", cioè informazioni accessibili solo da persone scelte dall'utente come possibili fruitori dei dati contenuti nel proprio profilo, con conseguente esclusione di tutti coloro che, iscritti o non iscritti, non sono stati autorizzati. Esse vengono nascoste tramite idonee misure e che l'utente decide di condividere esclusivamente con sé stesso o con una o più persone ben specificate. Ebbene, non presenta particolari problemi interpretativi la facoltà della polizia giudiziaria di avvalersi autonomamente, nel corso delle indagini, di tutti i dati pubblicamente accessibili in rete. Si tratta di una sorta di "pedinamento virtuale" che rientra in quell'attività atipica (artt. 55 e 348 del codice di rito) senza necessità di un provvedimento anticipato e autorizzativo della magistratura. Quanto ai dati riservati, la loro acquisizione ed il loro sfruttamento per fini investigativi non rientra nell'ambito dell'OSINT, che, per definizione, si avvale di fonti aperte. Ufficialmente, quindi, la polizia giudiziaria può ottenere questo tipo di informazioni esclusivamente attraverso la collaborazione del gestore del social.

Considerando le analisi al social più diffuso (Facebook), si osserva che le forze dell'ordine possono richiedere dati riservati attraverso le seguenti fondamentali modalità:

- Telematicamente (attraverso l'accesso al Request Secure Access to the Law Enforcement Online Request System)
- E-mail

- Fax
- Posta ordinaria

Il Facebook Security Law Enforcement Response Team evade le richieste pervenute sulla base di precisi criteri, «nel rispetto delle [. . .] condizioni di servizio e delle leggi applicabili, compreso il Federal Stored Communications Act ("SCA"), 18 U.S.C. Sezioni 2701-2712». In base alla legge statunitense, «Per procedere alla divulgazione di dati di base di un abbonato, è necessario un decreto ingiuntivo valido, rilasciato in connessione con un'investigazione di natura penale (18 U.S.C. Sezione 2703(c) [. . .].

Per obbligare Facebook a rivelare informazioni specifiche o altri dati relativi agli account, tra cui intestazioni di messaggi e indirizzi IP, in aggiunta ai dati essenziali sugli utenti, è necessaria un'ingiunzione del tribunale, come previsto dal Titolo 18 U.S.C., Articolo 2703. Sono invece esclusi i contenuti delle comunicazioni. Per obbligare Facebook a rivelare i contenuti memorizzati su un qualsiasi account, tra cui messaggi, foto, video, post in bacheca e informazioni sui luoghi, è necessario che, alla luce di "fondati motivi", venga emesso un mandato di perquisizione conforme alle procedure contenute nelle "Federal Rules of Criminal Procedure" o ad altre procedure statali sui mandati di perquisizione equivalenti». In Italia, "codice della privacy" alla mano, tali richieste di informazioni da parte delle forze dell'ordine possono considerarsi legittime nei limiti di seguito descritti:

- Richieste di informazioni formulate nell'ambito di attività di indagine di polizia giudiziaria
- Richieste avanzate da pubbliche autorità per altre finalità istituzionali

Le prime rientrano nei "trattamenti di dati personali" effettuati per «ragioni di giustizia» (art. 8, comma 2, lett. g, codice della privacy) ovvero «per finalità di prevenzione, accertamento o repressione di reati» (art. 53, comma 1, codice della privacy). In entrambi i casi, si deve dar corso a queste richieste purché sia chiaro il riferimento ad una attività di polizia giudiziaria, non ostacolando l'applicabilità del codice sulla privacy. Infatti, a norma dell'art. 132 del Codice in materia di protezione dei dati personali, «i dati relativi al traffico telefonico, sono conservati dal fornitore per ventiquattro mesi dalla data della comunicazione, per finalità di accertamento e repressione dei reati, mentre, per le medesime finalità, i dati relativi al traffico telematico, esclusi comunque i contenuti delle comunicazioni, sono conservati dal fornitore per dodici mesi dalla data della comunicazione». Tuttavia, riguardo il principio di pertinenza, per quanto possibile le informazioni divulgate dovranno essere circostanziate sotto il profilo oggettivo e temporale. Viceversa, se le richieste avanzate da forze di polizia o da altre pubbliche autorità non siano riconducibili all'esercizio di poteri di polizia giudiziaria, si applica la disciplina generale a tutela della privacy, in base alla quale un soggetto può permettere l'accesso a dati personali solo in adempimento di un obbligo legale, o, in alternativa, in presenza del consenso dell'interessato (cfr. artt. 23, comma 1, e 24, comma 1, lett. a, codice della privacy). Detto ciò, bisogna precisare che nella prassi operativa, queste informazioni vengono ottenute attraverso la "Sorveglianza online", ossia la tecnica che consente agli investigatori di rilevare e registrare da remoto ed in tempo reale tutto ciò che accade attraverso un determinato dispositivo (PC, tablet, smartphone, ecc.). Nel momento in cui il soggetto sfrutta questo dispositivo allo scopo di gestire online il proprio profilo registrato su un determinato social network, tale attività viene monitorata e "strappata".

5.5 Il Dark Web giuridicamente parlando

Con il termine Deep Web si fa riferimento al "web profondo", sommerso e oscuro che trova un'ulteriore sfumatura nel "Dark Web", ossia una sottoclasse che abbraccia un lato di internet inesplorato e che si spinge spesso oltre i limiti della legalità. A partire da alcuni anni, nel cyberspazio si parla del "Deep Web", raffigurato come minaccia alla sicurezza delle grandi architetture informatiche, tra cui quelle delle istituzioni politiche. Intorno a questi fenomeni sono apparse differenti definizioni le quali, attenuate spesso da un lessico prettamente giornalistico, sono sempre state accostate sullo sfondo di attività criminali. Non si può nascondere, tuttavia, che il Deep Web così come lo vediamo oggi sia invece nato con un intento meno "torbido" rispetto a come descritto nell'immaginario comune.

Nei Paesi colpiti dalle dittature militari, sottostanti a rigide politiche di censura della manifestazione del pensiero, la parte sommersa di internet, era infatti diventato l'unico spazio di comunicazione sicuro con il resto del mondo per giornalisti, associazioni di tutela dei diritti umani e dissidenti perseguitati dai governi. Il Deep Web era stato concepito come un luogo sicuro di scambio di informazioni in ambito commerciale ed industriale in quanto, grazie alla possibilità di crittografare la trasmissione di dati, era in grado di ridurre drasticamente il rischio di fuga dei segreti industriali. Con Deep Web siamo soliti riferirci a tutte quelle informazioni presenti su internet ma non ancora indicizzate dai più comuni e diffusi motori di ricerca (Google, Bing e Yahoo).

Vediamo ora, per precisazione, come in internet possiamo distinguere le pagine "statiche" dalle pagine "dinamiche":

- Le prime, contengono i comuni file con estensione .html che descrivono dettagliatamente testi da impaginare, grafica e immagini. Possiamo identificarle nelle pagine dei più comuni siti web il cui server, al momento del collegamento da parte dell'utente, invia al browser un file con estensione .html che viene decodificato per mostrare tutti i suoi contenuti

- Le seconde invece, non contengono file html bensì programmi per il server che vengono eseguiti direttamente dal browser. Solo in questo momento viene generato un codice html che non esisteva prima come nelle precedenti pagine statiche

Appare chiaro che una pagina dinamica ha potenzialmente la possibilità di generare molti codici html con informazioni diverse a seconda del caso, rendendo indeterminabile a priori il suo contento. Il Deep Web, pertanto, è costituito da pagine dinamiche così come definite sopra e ne fanno parte, ad esempio, le pagine web di nuova creazione, i web software, le reti private e le pagine indipendenti. Diversamente, con Dark Web si indica un vero e proprio sottoinsieme del Deep Web che contiene informazioni alle quali è possibile accedere pubblicamente, ma il cui accesso risulta essere più complicato dal fatto che l'indirizzo IP del dominio che ospita il sito risulta nascostosi trova su reti sovrapposte (overlay network). Esistono diverse architetture logiche con la quale possono essere distribuiti questi sistemi. Uno dei principali è il Peer-To-Peer (P2P), caratterizzato dall'assenza di gerarchia e dove ciascun nodo può essere a suo tempo sia Client sia Server. Un esempio di sistema P2P era Emule ("il mulo della felicità"), che ha rappresentato per anni il più grande sistema di file sharing al mondo. Nei sistemi P2P come Emule, infatti, non esiste un server centrale, ma lo scambio avviene direttamente tra gli utenti che scaricano e condividono nello stesso tempo frazioni dello stesso file. Tuttavia, qualora un utente voglia accedere al "mondo sommerso" di internet dovrà dotarsi di un particolare browser che gli permetta di navigare in totale anonimato tra i diversi nodi e livelli della rete. Il Dark Web si caratterizza per la presenza di architetture informatiche che vengono denominate "Darknet", ossia vere e proprie reti virtuali private che consentono l'interazione e lo scambio di informazioni tra gli utenti che ne fanno parte. Tra i più diffusi Darknet abbiamo "Tor" (acronimo di The Onion Router). Si tratta di un sistema di comunicazione per la navigazione anonima su internet che protegge l'utente dall'analisi del traffico e consente all'utente di falsificare il proprio indirizzo IP, garantendogli l'anonimato online. Il funzionamento della rete Tor è più semplice di quanto si possa pensare: mentre in un sistema tradizionale le informazioni transitano da un client

ad un server, in questo caso il percorso è frapposto dai server Tor che predispongono un vero e proprio circuito crittografato a strati. In base a questo, appare chiaro come lo spazio sommerso del Deep Web e soprattutto del Dark Web, si presti facilmente alla diffusione di attività illecite attraverso quelli che vengono denominati "black market". Sono portali multimediali di vendita e scambio di prodotti prevalentemente di origine e fattura illecita. Recentemente, una task force dell'FBI è riuscita a porre i sigilli a "Silk Road", una piattaforma commerciale delle più disparate tipologie di sostanze stupefacenti dotata di un sistema pressoché simile a quello del più noto (e lecito) sito di e-commerce Amazon.

Navigare nel deep web e/o nel dark web costituisce reato? In primo luogo, occorre osservare che accedere attraverso i vari portali al mondo sommerso del web, di per sé, non costituisce forma di reato. Il Codice penale qualifica come abusivi soltanto quegli accessi effettuati nei confronti di sistemi informatici o telematici protetti da misure di sicurezza ovvero contro la volontà espressa o tacita di chi ha il diritto di escluderlo (cfr. art. 615-ter c.p.). In questo modo quando si attua una potenziale condotta criminosa ci si sposta nella dinamica fattuale successiva, qualora l'utente attui ulteriori condotte prima della consumazione di reati disciplinati nel nostro ordinamento. Le contrattazioni nei black market, l'acquisto di beni provenienti da altro delitto, l'acquisto di sostanze stupefacenti, il download di materiale pedopornografico o l'acquisto di armi ed esplosivi, integreranno singolarmente le puntuali fattispecie di reato disciplinate nel nostro ordinamento.

6 Conclusione

6.1 Terza generazione di Osint

Nonostante si riconosca che esistono sempre più informazioni di valore di intelligence di pubblico dominio, l'IC è stato ancora lento ad abbracciare pienamente il potenziale dell'OSINT di seconda generazione. Mentre l'IC continua a confrontarsi su come gestire e sfruttare appieno l'OSINT di seconda generazione, è utile pensare alla prossima direzione del web e alle tendenze che potrebbero definire una terza generazione di OSINT. L'OSINT di seconda generazione si è evoluta in gran parte a causa del Web 2.0, uno spostamento del contesto di Internet verso pagine web dinamiche e contenuti generati dagli utenti.[102]

Tuttavia, per un decennio gli esperti di tecnologia hanno parlato dell'evoluzione verso il Web 3.0 - il "Web semantico" - che includerà nuove capacità di archiviazione ed elaborazione dei dati.

È inoltre probabile che la crittografia diventi una caratteristica più diffusa dell'OSINT di terza generazione, poiché il software di crittografia diventa sempre più pervasivo, accessibile e robusto. La decriptazione delle informazioni di potenziale valore per l'intelligence è generalmente di competenza dell'NSA.

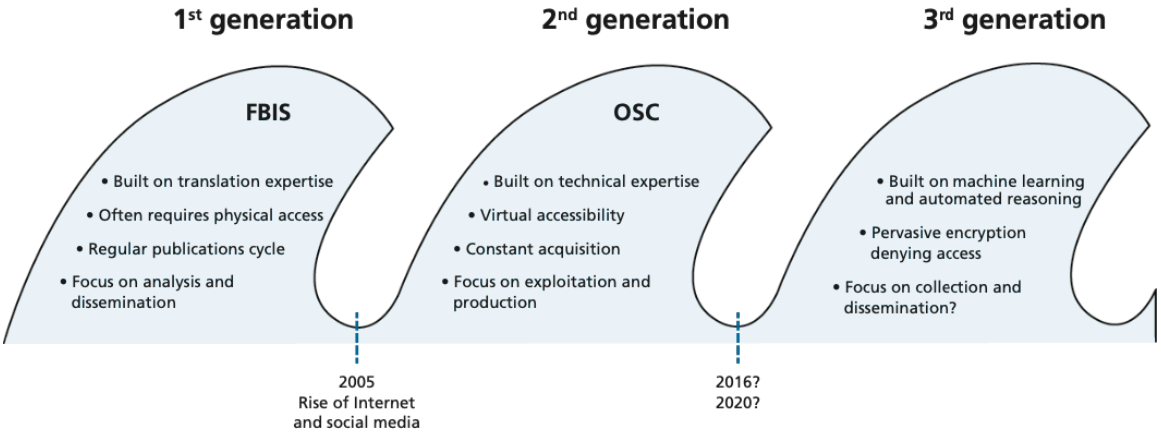


Figura 10: Caratteristiche delle generazioni OSINT[3]

6.2 Esempio contemporaneo dell'utilizzo dell'Osint: la guerra russo-ucraina

I social media hanno svolto un ruolo fondamentale nel fornire consapevolezza della situazione a milioni di persone dopo la guerra tra Russia e Ucraina, definita addirittura la "prima guerra di TikTok". Sebbene non sia una novità che i conflitti e i disordini si svolgano sui social media, l'evoluzione di funzioni come il live streaming e il fatto che oltre il 61% delle persone in Ucraina possieda uno smartphone[103] rendono i resoconti di prima mano di ciò che sta accadendo sul campo più accessibili rispetto alle guerre precedenti.

Circa il 61% delle persone in Ucraina possiede uno smartphone e, a differenza dei conflitti in Siria, è più facile per i media internazionali accedere ai luoghi del conflitto. Tuttavia, c'è la possibilità che l'enorme quantità di informazioni prodotte possa sopraffare gli investigatori, i media e il pubblico - in termini di volume e di violenza che potrebbero mostrare.

Il 9 marzo 2022, a meno di due settimane dall'inizio della guerra, i video su TikTok con l'hashtag "#ukraine" avevano già ottenuto complessivamente più di 26,8 miliardi di visualizzazioni.[7] Il conflitto è stato più popolare in termini di coinvolgimento dei contenuti su TikTok che su qualsiasi altra piattaforma.

[7] https://mediamanipulation.org/research/tiktok-war-ukraine-and-10-features-make-app-vulnerable-misinformation

Già prima dell'offensiva militare iniziata dalle forze armate russe, le prime avvisaglie dei preparativi dell'invasione ai confini dell'Ucraina sono stati notati attraverso il social network.

Nei primi mesi dell'anno, il Conflict Intelligence Team[8] (CIT) un'organizzazione investigativa indipendente con sede tra Russia e Ucraina che conduce indagini open-source sugli eventi che si verificano durante i conflitti armati, certifica un consistente spostamento di truppe russe verso il confine ucraino confrontando le informazioni registrate in un database centrale ferroviario con le immagini dei social media, ricercando alcuni hashtag di TikTok in lingua russa, all'epoca abbondanti di video di treni che trasportavano sistemi missilistici a lancio multiplo, porta truppe e carri armati.[9] Grazie a queste immagini i loro ricercatori hanno abbinato visivamente le stazioni al percorso dei treni.

Secondo Kirill Mikhailov, uno dei componenti del CIT, in alcuni casi il tipo di hardware presente su un treno può essere abbinato a specifiche formazioni militari. Il CIT e altri ricercatori hanno individuato nei video dei social media attrezzature presumibilmente utilizzate da unità della storica 76a divisione paracadutista russa Guards Air Assault, ad esempio, grazie agli specifici veicoli utilizzati, alle verniciature distintive o ai contrassegni dell'unità. Questo tipo di informazioni può essere incrociato con la base nota di un'unità militare.[104][10]

I ricercatori del CIT integrano le loro scoperte anche con dati satellitari o, in alcuni casi, con commenti sui social media. Qualora un video su TikTok diventasse particolarmente virale, potrebbe attirare i commenti dei parenti dei soldati, contenenti informazioni utili come il suggerimento che la missione militare del proprio caro sarà più lunga delle esercitazioni di routine annunciate pubblicamente dall'esercito russo.[11]

Per evitare disinformazione, Mikhailov ha detto che i ricercatori della CIT mirano a raccogliere i post sui social media di veri testimoni oculari. Anche le immagini satellitari aiutano a verificare i dati raccolti da CIT, ma un metodo utilizzato da CIT per convalidare le proprie scoperte è diventato di recente più complicato, dopo un intervento delle autorità russe. I numeri a otto cifre riportati sulla fiancata di un vagone possono aiutare il CIT a isolare un treno specifico e a ottenere una cronologia dei suoi movimenti. Ora è più difficile ottenere questi dati, ha detto Mikhailov.

Sempre attorno il mese di gennaio, un'altra organizzazione, il Centre for Information Resilience (CIR), un'organizzazione no-profit indipendente che si occupa di contrastare la disinformazione, denunciare le violazioni dei diritti umani e combattere i comportamenti online dannosi per le donne e le minoranze,[105] ha mappato e verificato video ripresi dalla popolazione locale che tracciano i movimenti di attrezzature militari e truppe lungo i fianchi orientali dell'Ucraina.[12] I video girati nei dintorni della città russa di Kursk, a circa 100 miglia dal confine con l'Ucraina, mostrano auto in fila per attraversare i binari dei treni utilizzati per trasportare i carri armati da un luogo all'altro. Decine di veicoli militari sono stati filmati parcheggiati insieme.[13] Filmati traballanti mostrano carri armati che si muovono su un terreno innevato lungo una strada trafficata.[14]

Sin dall'aprile 2021, la mobilitazione delle truppe russe è stata accompagnata da numerose prove digitali. Queste provengono da una varietà di fonti, dai filmati degli smartphone alle immagini dall'alto ad alta risoluzione catturate da società satellitari commerciali. Truppe, elicotteri e attrezzature militari sono stati avvistati nelle immagini satellitari. Ma per le persone sul campo, TikTok è emerso come una piattaforma chiave per mostrare i movimenti militari.

[8]La ricerca del CIT è stata ampiamente citata negli ultimi mesi, anche in un'analisi del 15 gennaio sull'accumulo militare della Russia nei pressi dell'Ucraina da parte di due esperti dell'organizzazione no-profit CNA, un think tank che fornisce consulenza alle forze armate statunitensi. Il lavoro del gruppo è apparso anche nelle recenti pubblicazioni del Digital Forensic Research Lab del Consiglio Atlantico.

[9]`https://twitter.com/CITeam_en/status/1485997570921009166`

[10]`https://twitter.com/CITeam_en/status/1381615350618460166`

[11]`https://twitter.com/CITeam_en/status/1483135533635227654`

[12]https://maphub.net/Cen4infoRes/russian-ukraine-monitor

[13]https://twitter.com/GirkinGirkin/status/1492522158282747908

[14]https://twitter.com/4emberlen/status/1491418129414914048

Secondo Eliot Higgins, fondatore del collettivo internazionale indipendente open source Bellingcat[106], che da anni denuncia lo spionaggio russo, "TikTok è sicuramente una delle principali piattaforme utilizzate per documentare questo fenomeno".

I video di TikTok intorno a Kursk - la cui posizione è stata verificata dal CIR - forniscono un'istantanea di quanto sia diventata potente l'intelligence open source. I video contribuiscono ai rapporti dei media e alle discussioni politiche, possono essere di bassa qualità grafica, ma mostrano esattamente ciò che sta accadendo in un momento specifico.

Secondo quanto riporta Wired, gli account che condividono brevi filmati di rifornimenti di truppe spesso sembrano essere stati postati da persone normali: video di carri armati che si muovono accanto a filmati di bambini che giocano. Questo significa anche che non sono in molti a guardarli. Molti video verificati intorno a Kursk hanno meno di 1.000 visualizzazioni e ancora meno commenti e condivisioni.

Mentre alcuni condividono semplicemente video di qualcosa di insolito che sta accadendo nella loro città, altri usano TikTok in modo molto più deliberato. Strick afferma di aver visto numerosi account TikTok anonimi creati per caricare e condividere filmati di attività militari russe intorno al confine ucraino.[103] Per continuare a ricevere lo stesso tipo di contenuti, il CIR ha creato nuovi account TikTok per addestrare il suo algoritmo a mostrare filmati di movimenti militari intorno alla Russia. L'algoritmo di TikTok, sebbene poco trasparente, raccomanda video simili a quelli che le persone hanno già guardato, apprezzato o seguito. Nel giro di poche settimane, il CIR ha addestrato l'algoritmo di TikTok a mostrare un flusso costante di video che sembrano provenire dalla regione.

6.2.1 Propaganda russa

Nel mondo delle indagini open source, la velocità e l'accuratezza hanno grande importanza. I video e le immagini che circolano online devono essere verificati rapidamente per garantire che le informazioni false non abbiano l'opportunità di plasmare le narrazioni. I giornalisti e chi posta sui social media coprono gli eventi in diretta e possono commettere errori subito dopo l'accaduto. Se da una parte questa abbondanza di informazioni può aiutare a dipingere un quadro più dettagliato del contesto militare attorno alla guerra, dall'altro può diventare una trappola al servizio della propaganda avversaria se le fonti non vengono debitamente verificate. Poco prima della invasione in Ucraina, sui social media e sui media statali russi sono apparsi una serie di video che sembravano suggerire un'aggressione ucraina nei pressi del confine con la Russia e di due repubbliche autoproclamate (regioni occupate riconosciute in modo controverso dalla Russia all'inizio di questa settimana) nella parte orientale del Paese.

I video in questione hanno accumulato innumerevoli visualizzazioni online, ma a un esame più attento e utilizzando le risorse open source, il collettivo investigativo Bellingcat ha dimostrato come molti di questi video facciano parte di una propaganda fabbricata per diffamare o demoralizzare il governo e le forze armate dell'Ucraina che devono affrontare l'attacco della Russia. Uno dei primi video sospetti emersi durante le crescenti tensioni tra Ucraina e Russia, pubblicato il 18 febbraio, mostra presumibilmente membri della regione separatista ucraina della Repubblica Popolare di Donetsk, che affrontano sabotatori ucraini di lingua polacca il cui obiettivo era far esplodere un serbatoio di cloro nel territorio controllato dai separatisti.[107] L'incidente è stato riportato dall'agenzia di stampa statale russa Tass[108] e condiviso dai circoli filo-russi sui social media come un dato di fatto, sostenendo che il filmato è fosse stato recuperato dal corpo di uno dei sabotatori. Tuttavia, la comunità online ha iniziato a individuare rapidamente una serie di problemi con le prove presentate dall'agenzia. L'utente @oldLentach ha infatti esaminato i metadati del video e trovato una data di creazione del video risalente all'8 febbraio, dieci giorni prima della presunta data dello scontro a fuoco.[109] Nei metadati era presente anche un nome di file, "M72A5 LAW and AIPLAS live fire.mp4" nella sezione "Pantry" dei metadati. Questa contiene i dettagli di altri file che compongono il file pubblicato, come altre fonti di audio e video. Ciò significa che al file video sono stati aggiunti altri file audio o video. Una rapida ricerca su YouTube rivela un solo video con lo stesso nome del file nei metadati del video del "sabotatore polacco", una

registrazione del 2010 di una esercitazione militare delle Forze di Difesa Finlandesi, contenente una serie di scoppi ed esplosioni.[15] Le forme d'onda dell'audio specifiche di alcune esplosioni del video su YouTube corrispondono a quelle del video dei "sabotatori polacchi".

Il governo russo ha rivendicato l'aggressione ucraina in alcuni altri report, tra cui un filmato, rilasciato il 18 febbraio, presumibilmente registrato da una videocamera posizionata sull'elmetto di un soldato ucraino intento in un'altra operazione di sabotaggio nell'Oblast' di Rostov, al confine tra Ucraina e Russia. I media russi hanno in seguito riportato l'uccisione di cinque soldati ucraini nell'operazione e di aver distrutto due veicoli blindati ucraini.[110] Giornalisti russi si sono recati poi sul posto dopo il fatto[111] mostrando da più vicino i veicoli. Poche ore dopo la pubblicazione del video, (al di là della perplessità generata dal voler attribuire la produzione di un filmato così incriminante a un soldato in un'operazione di sabotaggio) è stata fatta notare la scarsa verosimiglianza dell'attacco, in quanto il luogo dello scontro era in quel momento ben addentro i territori occupati dai separatisti filorussi.[16] Ciò significa che i militari ucraini avrebbero dovuto attraversare circa 40 chilometri di territorio nemico prima di raggiungere il confine con la Russia. Gli incendi causati dalle esplosioni dei veicoli blindati sono persino apparsi sul Fire Information for Resource Management System della NASA, che utilizza i satelliti per tracciare gli incendi in tutto il mondo.[112] A certificare che questa si tratti di una operazione di false flag[17] è la questione sul tipo di veicoli distrutti. I veicoli utilizzati in questa cosiddetta incursione sono BTR 70 M, un tipo di veicolo corazzato per il trasporto di personale che non utilizzato dalle forze ucraine. Tra l'altro, il luogo in cui è stato girato questo video è lo stesso in cui la Russia ha dichiarato che uno dei suoi posti di frontiera è stato distrutto dai bombardamenti ucraini.

[15] https://www.youtube.com/watch?v=5T3Oc3iuSO8

[16] `https://twitter.com/EliotHiggins/status/1495775073906610180`

[17] Una operazione false flag è una qualsiasi operazione commessa con l'intento di mascherare l'effettiva fonte di responsabilità e incolparne un'altra. Consiste in una tattica segreta perseguita con operazioni militari, attività di intelligence e/o spionaggio, condotte in genere da governi, servizi segreti, progettata per apparire come perseguita da altri enti e organizzazioni[113]

7 Tabelle Tools

7.1 Tabella Tools Social Network

Nome tool	Ultimo commit	Funzionalità	Descrizione	OS	Prezzo	Rating
accountanalysis	Sconosciuto	SN	Valuta l'attività degli account Twitter.	WA	G/P	⋆⋆
Epieos	Sconosciuto	IG/SN	Recupero informazioni legate a indiriz email.	WA	G/P	⋆ ⋆ ⋆
Digital image Forensic	Sconosciuto	IA	Analisi forense di immagini, la versio web di Ghiro.	WA	G	⋆⋆
Ghiro	16/09/2016	IA	Esegue analisi forensi su più immagini	L	G	⋆ ⋆ ⋆
GHunt	04/07/2022	IG/SN	Information gathering partendo da em e documenti Google	L	G	
GvngSearch	19/03/2022	SN	Ricerca informazioni personali trami social media	L	G	⋆
have i been pwned?	Sconosciuto	IG	Verifica se i dati associati a mail sono st compromessi	WA	G	⋆ ⋆ ⋆
Holehe	21/07/2022	IG	Analisi email	L	G	⋆ ⋆ ⋆
Ignorant	27/06/2021	IG	Controlla se un numero di telefono è c legato ad account Social	L	G	⋆
Lyzem	Sconosciuto	IG	Telegram search engine per canali, grup ecc	WA	G	⋆
Mosint	11/06/2022	IG	Analisi Email	L	G	⋆ ⋆ ⋆
Mr. Holmes	01/09/2022	IG	Ottiene info su domini, nomi utente telefono	L	G	⋆ ⋆ ⋆ ⋆ ⋆
Octosuite	14/02/2022	IG	Ricerca informazioni utilizzando Git H	L	G	⋆ ⋆ ⋆
Osintgram	05/09/2022	SN	Analisi su account Instagram	L	G	⋆⋆
PimEyes	Sconosciuto	IA	Ricerca inversa delle immagini basata s riconoscimento facciale.	WA	G/P	⋆ ⋆ ⋆
Quidam	18/05/2022	IG	Recupero informazioni su account ins gram e twitter	L	G	⋆
recon-ng	25/08/2021	IG	IG automatizzato.	L	G	⋆ ⋆ ⋆⋆
Sherlock	26/09/2022	IG	Ricerca accounts collegati a un deterr nato username	L	G	⋆ ⋆ ⋆⋆
Sherloq	06/08/2022	IA	Analisi delle immagini	L	G	⋆ ⋆ ⋆⋆
Snap-Scraper	30/08/2022	G	Scarica media da Snapchat in base a ge localizzazione	M	G	⋆
Spiderfoot	26/09/2022	IG	Ottiene info su domini, nomi utente telefono	L	G	⋆ ⋆ ⋆ ⋆ ⋆
Terra	28/05/2022	SN	Ricercati dati utenti Instagram/Twitter.	L	G	
theHarvester	27/09/2022	IG	Raccolta di informazioni di per pentesti	L	G	⋆ ⋆ ⋆⋆
TinEye	Sconosciuto	IA	Strumento per ricercare dove le foto ap iono online	WA	G/P	⋆ ⋆ ⋆
Toutatis	05/06/2022	SN/G	Analisi account Instagram.	L	G	⋆⋆
Twayback	24/08/2022	IG	Scarica i Tweet cancellati archiviati d l'utente target	W, L	G	⋆⋆
Twint	02/03/2021	SN	Ricerca dati nei tweet di un utente.	L	G	

Tabella 1: Legenda

OS	Prezzo	Funzionalità
W=Windows	G = Gratuito	IG = Information Gathering
L=Linux	P = A pagamento	IA = Image Analysis
M=Mac		G = Geolocalizzazione
WA=Web App		SN = Analisi Social network

7.2 Tabella Tools Geolocalizzazione

Nome tool	Ultimo commit	Funzionalità	Descrizione	OS	Prezzo	Rating
Echosec	Sconosciuto	G	Individuazione di tweet, post e persone una determinata area	WA	P	⋆⋆
Foto forensics	Sconosciuto	IA	Analisi delle immagini meta e specifici	WA	G	⋆ ⋆ ⋆⋆
Geolocation Estimation	Sconosciuto	IG	Geolocalizzazione usando immagini	WA	G	⋆ ⋆ ⋆
GeoWifi	02/09/2022	G	Geolocalizzazione di reti wifi usando B SID e SSID	L	G	⋆⋆
Live Ua map	Sconosciuto	EG	Geolocalizzazione delle guerre, crimi ed altri eventi nel mondo	WA	G	⋆ ⋆ ⋆⋆
Map Hub	Sconosciuto	SNG	Geolocalizzazione dei vari post riguar la guerra in ucraina	WA	G	⋆ ⋆ ⋆⋆
SunCalc	Sconosciuto	CL	Geolocalizzazione temporale utilizzan sistema delle ombre e posizione del so	WA	G	⋆ ⋆ ⋆
Telegram nearby map	02/03/2022	TG	Geolocalizzazione utenti connessi a Te gram.	W	G	⋆ ⋆ ⋆⋆
Youtube-geofind	Sconosciuto	YTG	Analisi forense di immagini, la versio web di Ghiro.	W/L	G	⋆ ⋆ ⋆⋆
What is where	Sconosciuto	G	Ricerche tramite query su OpenStreetM di POI (point of interest)	WA	G	⋆ ⋆ ⋆⋆
Yandex	Sconosciuto	G	Geolocalizzazione di un luogo attraver reverse image	WA	G	⋆ ⋆ ⋆
Google dork	Sconosciuto	IG	Affinazione ricerche Google	EF	G	⋆ ⋆ ⋆
Google Images	Sconosciuto	IA	Analisi di immagini, riconoscimento associazione ad immagini simili	WA	G	⋆ ⋆ ⋆⋆
Sherloq	06/08/2022	GM	Analisi delle immagini	L	G	⋆ ⋆ ⋆
BatchGeo	Sconosciuto	GM	Creazione mappa e inserimento luog analizzati e mappatura	WA	G	⋆ ⋆ ⋆
PimEyes	Sconosciuto	IA	Ricerca inversa delle immagini basata s riconoscimento facciale.	WA	G/P	⋆ ⋆ ⋆
Pic2Map	Sconosciuto	IA	Geolocalizzazione di un foto e visualiz zione di dati exif	WA	G	⋆ ⋆ ⋆
World Webcams	Sconosciuto	WC	Connessione a tutte le webcam disponib in giro per il mondo	WA	G	⋆⋆
Map Channels	Sconosciuto	M	Permette di visualizzare la posizione c la streetview, satellite e mappa normal	WA	G	⋆ ⋆ ⋆⋆
Gps Visualizer	Sconosciuto	GM	Permette di mappare le posizione inser do una sheet, file csv o google drive	WA	G	⋆ ⋆ ⋆⋆
Ads-B Exchange	Sconosciuto	T	Permette il tracking mondiale di veic volanti (aerei, elicotteri, droni etc)	WA	G	⋆ ⋆ ⋆
Google My Sheet	Sconosciuto	GM	Mappatura posizioni attraverso una she	WA	G	⋆ ⋆ ⋆
Gps Jam	Sconosciuto	IG	Visualizzazione zone nel mondo dov segnali sembrano essere distrubati	WA	G	⋆ ⋆ ⋆
Carnet	Sconosciuto	RI	Metodo per trovare le informazioni att verso la foto di un'auto	WA	G	⋆ ⋆ ⋆
Rare earth element	Sconosciuto	GM	Localizzazione nella mappa di tutti i lu ghi con depositi di RRE	WA	G	⋆⋆
TinEye	Sconosciuto	IA	Strumento per ricercare dove le foto ap iono online	WA	G/P	⋆ ⋆ ⋆
Wigle	Sconosciuto	GM	Localizzazione nella mappa dei disposi vi connessi ad una rete wifi	WA	G	⋆ ⋆ ⋆⋆
Youtube metadata	Sconosciuto	VA	Ricerca meta dati in un video pubblica su YouTube	WA	G	⋆ ⋆ ⋆
Url Scan	Sconosciuto	US	Scannerizzatore di url di foto con resti zione info e certificati	WA	G	⋆ ⋆ ⋆
Wolfaramalpha	Sconosciuto	CI	Processamento query , db computaziona	WA	G	⋆ ⋆ ⋆⋆

Tabella 2: Legenda

OS	Prezzo	Funzionalità
W=Windows	G = Gratuito	IG = Information Gathering
L=Linux	P = A pagamento	IA = Image Analysis
M=Mac		G = Geolocalizzazione
WA=Web App		TG = Telegram geolocalization
EF=Estensione Firefox		YTG= Youtube Geolocalization
		CL = Chrono-localization
		SNG= Social network geolocalization
		M = Maps
		GM = Geospatial mapping
		VA = Video analysis
		US = Url Scanner
		CI = Computational intelligence
		IG = Image geolocalization
		EG = Events geolocalization
		T = Tracking

Riferimenti bibliografici

[1] Office of the Director of Intelligence. What is intelligence, 2022. URL https://www.dni.gov/index.php/what-we-do/what-is-intelligence.

[2] Office of the Director of National Intelligence. U.s. national intelligence: An overview 2011, 2011.

[3] Heather J. Williams and Ilana Blum. *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*. RAND Corporation, Santa Monica, CA, 2018. doi: 10.7249/RR1964.

[4] Cameron Colquhoun. A brief history of open source intelligence, Luglio 2016. URL https://www.bellingcat.com/resources/articles/2016/07/14/a-brief-history-of-open-source-intelligence/.

[5] Central Intelligence Agency. Establishment of the dni open source center, 2005.

[6] Tim O'Reilly. What is web 2.0: Design patterns and business models for the next generation of software. https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html, 2007.

[7] Records of the foreign broadcast intelligence service, 2022. URL https://www.archives.gov/research/guide-fed-records/groups/262.html.

[8] Joseph E. Roop. *Foreign Broadcast Information Service History. Part I: 1941–1947*. Central Intelligence Agency, Washington, D.C., Aprile 1969.

[9] William J. Donovan. Intelligence: Key to defence. *Life*, 21(14):114, Sep 1946. https://books.google.co.uk/books?id=akkEAAAAMBAJ.

[10] Brian Rotheray. *A History of BBC Monitoring*. BBC Monitoring, Caversham Park, Reading, Regno Unito, 2009. URL https://web.archive.org/web/20110911081444/http://www.monitor.bbc.co.uk/about_us/BBCMhistory%20revisions%20x.pdf.

[11] *BBC Handbook*. British Broadcasting Corporation, 1940. URL https://worldradiohistory.com/UK/BBC/BBC-Annual/BBC-Year-Book-1940.pdf.

[12] Stephen C. Mercado. Fbis against the axis, 1941-1945. *Studies in Intelligence*, 45 (5), 2001. URL https://www.cia.gov/static/96048eae9f1b9aa309a24c4b5582ea62/fbis-against-the-axis.pdf.

[13] Deane J.Allen e Brian G. Shellum. *Defense Intelligence Agency: At the Creation, 1961–1965*. Defense Intelligence Agency, January 2002.

[14] J. Niles Riddel. *Remarks at the First International Symposium, 'National Security and National Competitiveness: Open Source Solutions'*. Open Source Solutions, Inc., January 2002.

[15] Harold P. Ford. Calling the sino-soviet split. *Studies in Intelligence*, 42(5):57–71, Inverno 1998–99. URL https://www.cia.gov/static/0884d1ea5f58fcaec06584742ed16442/Calling-Sino-Soviet-Split.pdf.

[16] J. J. Bagnall. The exploitation of russian scientific literature for intelligence purposes. *Studies in Intelligence*, 2(3):45–49, 1958. URL https://www.cia.gov/static/2b3f0edd9d643a76ce74e16a38af0931/Exploitation-Russian-Scientifc-Literature.pdf.

[17] Davis W. Moore. Open sources on soviet military affairs. *Studies in Intelligence*, 7:101, 1963. URL https://www.cia.gov/readingroom/docs/DOC_0000608367.pdf.

[18] Herman L. Croom. The exploitation of foreign open sources. *Studies in Intelligence*, 13:129–30, 1969. URL https://nsarchive2.gwu.edu/NSAEBB/NSAEBB90/dubious-07b.pdf.

[19] John O. Koehler. *Stasi: The Untold Story of the East German Secret Police*. Westview Press Boulder, Colo, 1999.

[20] Antonio J. Mendez. *The master of disguise*. William Morrow & Company, 1999.

[21] William O. Studeman. Teaching the giant to dance: Contradictions and opportunities in open source information. *Competitive Intelligence Review*, 4(1):25–29, 1992. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cir.3880040106.

[22] Ben Barber. "cia media translations may be cut: Users rush to save valuable resource". *Washington Times*, 30 Dicembre 1996.

[23] Nahid Siamdoust. Tehran's rallying cry: 'we are the people of iran', Jun 2009. URL https://web.archive.org/web/20090618140519/http://www.time.com/time/world/article/0,8599,1904764,00.html?xid=rss-topstories.

[24] Nic Newman. *The rise of social media and its impact on mainstream journalism (Publisher's version, Reuters Institute for the Study of Journalism: Working Papers)*. Reuters Institute for the Study of Journalism, 2016. URL http://www.sssup.it/UploadDocs/6635_8_S_The_rise_of_Social_Media_and_its_Impact_on_mainstream_journalism_Newman_07.pdf.

[25] Iran internet stats and telecommunications reports, Apr 2018. URL https://www.internetworldstats.com/me/ir.htm.

[26] Iran media stats, 2009. URL https://www.nationmaster.com/country-info/profiles/Iran/Media#2008.

[27] Internet brings events in iran to life, Jun 2009. URL http://news.bbc.co.uk/2/hi/middle_east/8099579.stm.

[28] Evgeny Morozov. Iran elections: A twitter revolution?, Jun 2009. URL https://www.washingtonpost.com/wp-dyn/content/discussion/2009/06/17/DI2009061702232.html.

[29] Martin Pasquier. The iran mobile market: Connectivity in 2014, Jul 2014. URL http://innovationiseverywhere.com/iran-mobile-market-connectivity-in-2014/.

[30] Joseph Fitsanakis. Analysis: Cia open source center monitors facebook, twitter, blogs, Nov 2011. URL https://intelnews.org/2011/11/08/01-861/#more-7508.

[31] David S. Cohen. Analysis: Cia open source center monitors facebook, twitter, blogs, Settembre 2015. URL https://web.archive.org/web/20200807114834/https://www.cia.gov/news-information/speeches-testimony/2015-speeches-testimony/deputy-director-cohen-delivers-remarks-on-cia-of-the-future-at-cornell-university.html/.

[32] Gordon Corera. The spies of tomorrow will need to love data, Aprile 2016. URL https://www.wired.co.uk/article/spies-data-mi6-cia-gordon-corera.

[33] Mark M. Lowenthal. Osint: The state of the art, the artless state. *Studies in Intelligence*, 45(3):63, 2001.

[34] Stephen C. Mercado. Sailing the sea of osint in the information age. *Studies in Intelligence*, 48(3): 45–55, 2004.

[35] Public law 109-163, 2006.

[36] Loch .K. Johnson. *Handbook of Intelligence Studies*, page 132. Routledge, New York, 2007.

[37] Nancy Cameron, Darla e Scola. Mapping the world's 4.3 billion internet addresses. *Washington Post*, Gennaio 2015.

[38] Intelligence community directive number 301: National open source enterprice, Luglio 2006.

[39] R.Paul Soule, Mason H. e Ryan. *Gray Literature*. Defense Technical Information Center, 1995.

[40] *Intelligence Guide for First Responders*. Joint Chiefs of Staff, 2013.

[41] Alfred Best, Richard Jr. e Cumming. Open source intelligence (osint): Issues for congress. *Issues for Congress*, December 2007.

[42] Libor Benes. OSINT, new technologies, education: Expanding opportunities and threats. a new paradigm. *Journal of Strategic Security*, 6(3Suppl):22–37, September 2013. doi: 10.5038/1944-0472.6.3s.3. URL `https://doi.org/10.5038/1944-0472.6.3s.3`.

[43] U.s. national intelligence — an overview 2011, 2011. URL `https://www.dni.gov/files/documents/IC_Consumers_Guide_2011.pdf`.

[44] Helen Gibson. Acquisition and preparation of data for osint investigations. In P. Saskia Bayerl e Fraser Sampson Babak Akhgar, editor, *Open source intelligence investigation: From Strategy to Implementation*, pages 69–93. Springer, Basilea, Svizzera, 2016.

[45] Nihad A. Hassan e Rami Hijazi. *Open source intelligence methods and tools – A practical guide to Online intelligence*. APress, Berkeley, 2018.

[46] Michael Bazzell. *Open source intelligence techniques: resources for searching and analyzing Online information*. Createspace Independent Publishing Platform, North Charleston, SC, 2016.

[47] Javier Pastor-Galindo, Pantaleone Nespoli, Felix Gomez Marmol, and Gregorio Martinez Perez. The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends. *IEEE Access*, 8: 10282–10304, 2020. doi: 10.1109/access.2020.2965257. URL `https://doi.org/10.1109/access.2020.2965257`.

[48] Fahimeh Tabatabaei e Douglas Wells. Osint in the context of cyber-security. In P. Saskia Bayerl e Fraser Sampson Babak Akhgar, editor, *Open source intelligence investigation: From Strategy to Implementation*, pages 213–231. Springer, Basilea, Svizzera, 2016.

[49] Arthur S. Hulnick. The dilemma of open source intelligence: Is osint really intelligence? In Loch K. Johnson, editor, *The Oxford Handbook of National Security Intelligence*. Oxford University Press, Oxford, March 2010. doi: 10.1093/oxfordhb/9780195375886.001.0001. URL `https://doi.org/10.1093/oxfordhb/9780195375886.001.0001`.

[50] Samuel Jacoby e Damien Cave Timothy Williams, James Thomas. Police body cameras: What do you see? *New York Times*, Apr 2016. URL `https://www.nytimes.com/interactive/2016/04/01/us/police-bodycam-video.html`.

[51] Noah Shachtman. Open source intel rocks—sorry, it's classified. *Wired Magazine*, Settembre 2008. URL https://www.wired.com/2008/09/download-hayden/.

[52] Isabella Böhm e Samuel Lolagar. Open source intelligence introduction, legal, and ethical considerations. *International Cybersecurity Law Review*, 2(2):317–337, nov 2021. doi: 10.1365/s43439-021-00042-7. URL https://doi.org/10.1365/s43439-021-00042-7.

[53] Brian Everstine. Carlisle: Air force intel uses isis 'moron's' social media posts to target airstrikes, Giugno 2015. URL https://www.airforcetimes.com/news/your-air-force/2015/06/04/carlisle-air-force-intel-uses-isis-moron-s-social-media-posts-to-target-airstrikes/.

[54] Bellingcat. Identifying the separatists linked to the downing of mh17, 2019. URL https://www.bellingcat.com/news/uk-and-europe/2019/06/19/identifying-the-separatists-linked-to-the-downing-of-mh17/.

[55] Bellingcat. Full report: Skripal poisoning suspect dr. alexander mishkin, hero of russia, 2018. URL https://www.bellingcat.com/news/uk-and-europe/2018/10/09/full-report-skripal-poisoning-suspect-dr-alexander-mishkin-hero-russia/.

[56] Eliot Higgins. All the pieces matter - syria's chlorine bombs and the douma chemical attack, 2018. URL https://www.bellingcat.com/news/mena/2018/04/29/pieces-matter-syrias-chlorine-bombs-douma-chemical-attack/.

[57] Adam Rawnsley. Syria's "new" iranian drone, 2016. URL https://www.bellingcat.com/news/mena/2016/01/28/syria-new-iranian-drone/.

[58] Oleksiy Kuzmenko. The curious case of david jewberg, the fake senior pentagon russia analyst, 2018. URL https://www.bellingcat.com/news/americas/2018/04/02/curious-case-david-jewberg-fake-senior-pentagon-russia-analyst/.

[59] Syrian Archive. Methodology: How we tracked the illegal shipment of sarin precursor from belgium to syria, 2018. URL https://www.bellingcat.com/resources/case-studies/2018/04/19/methodology-tracked-illegal-shipment-sarin-precursor-belgium-syria/.

[60] Günther Epple e Franziska Ludewig. Open source intelligence in einsatzleitstellen der polizei: eine empirische untersuchung zu neuen möglichkeiten der informationsgewinnung. *Schriftenreihe der Deutschen Hochschule der Polizei*, 11, 2020.

[61] Steve Ramwell e Tony Day Helen Gibson. Analysis, interpretation and validation of open source data. In P. Saskia Bayerl e Fraser Sampson Babak Akhgar, editor, *Open source intelligence investigation: From Strategy to Implementation*, pages 95–110. Springer, Basilea, Svizzera, 2016.

[62] Kevin D. Mitnick e William L. Simon. *L'arte dell'inganno. I consigli dell'hacker più famoso del mondo*. Feltrinelli, 2003. ISBN 9788807170867.

[63] Christina Lekati e Samuel Lolagar. Why for today's cyber investigations we need to combine intelligence disciplines, May 2021. URL https://christina-lekati.medium.com/why-for-todays-cyber-investigations-we-need-to-combine-intelligence-disciplines-afca5363048c

[64] Europol. Stop child abuse – trace an object, 2017. URL https://www.europol.europa.eu/stopchildabuse.

[65] Meredith May. Vast search off coast for data wizard, Jan 2007. URL https://www.sfgate.com/news/article/Vast-search-off-coast-for-data-wizard-2620302.php.

[66] Joseph M. Hellerstein e David L. Tennenhouse. Searching for jim gray: A technical overview. *Commun. ACM*, 54(7):77–87, jul 2011.

[67] Ramian Fathi. Soziale medien in katastrophen - herausforderungen und lösungsansätze in einer hochvernetzten gesellschaft. *BUW-Output*, 25:24 –29, Luglio 2021.

[68] Tom Smith Quentin Revell and Robert Stacey. Tools for osint-based investigations. In P. Saskia Bayerl e Fraser Sampson Babak Akhgar, editor, *Open source intelligence investigation: From Strategy to Implementation*, pages 153–166. Springer, Basilea, Svizzera, 2016.

[69] Cyber risk platform, 2022. URL https://blackkite.com/platform/.

[70] Kim kardashian west robbed of millions by paris gunmen, Ottobre 2016. URL https://www.bbc.com/news/world-europe-37538453.

[71] Kenzie Bryant. Kim kardashian's alleged robber confirms social media helped him plan heist, Gennaio 2017. URL https://www.vanityfair.com/style/2017/01/kim-kardashian-paris-robbery-social-media-heist.

[72] Jordan Valinsky. Topsy, the internet's favorite social media analysis tool, has died at 8. *Digiday*, 2015. URL https://digiday.com/marketing/topsy-the-internets-favorite-social-media-analysis-tool-has-died-at-8/.

[73] Daisuke Wakabayashi e Douglas MacMillan. Apple taps into twitter, buying social analytics firm topsy. *Wall Street Journal*, 2013. URL https://www.wsj.com/amp/articles/DJFVW00020131202e9c2sq2om.

[74] Alex Hern. "google says machine learning is the future. so i tried it myself". *The Guardian*, 2016. URL https://www.theguardian.com/technology/2016/jun/28/google-says-machine-learning-is-the-future-so-i-tried-it-myself.

[75] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 1994. doi: 10.1017/CBO9780511815478.

[76] Robert A. Hanneman and Mark Riddle. *Introduction to Social Network Methods*. University of California, Gennaio 2005.

[77] C.A. Heaney and B.A. Israel. Social networks and social support. *Health Behavior and Health Education: Theory, Research, and Practice*, 3:189–210, 01 2008.

[78] Linton C. Freeman. *Centrality in Social Networks: Conceptual Clarification*, volume 1, chapter 215–239. Social Networks, 1978.

[79] Rami Puzis, Dana Yagil, Yuval Elovici, and Dan Braha. Collaborative attack on internet users' anonymity. *Internet Research*, 19:60–77, 2009.

[80] Chiara Livia Bernardi. *Digital media and women's issues in Egypt and Saudi Arabia*. PhD thesis, University of Warwick, Warwick, 2015.

[81] Elizabeth Bodine-Baron, Todd C. Helmus, Madeline Magnuson, and Zev Winkelman. *Examining ISIS Support and Opposition Networks on Twitter*. RAND Corporation, Santa Monica, CA, 2016. doi: 10.7249/RR1328.

[82] Pete Burnap e Matthew Williams Luke Sloan, Jeffrey Morgan. Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. *PLoS One*, 10(3), mar 2015.

[83] Ewa Witalisz and Justyna Leśniewska. Native vs. non-native english: data-driven lexical analysis. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 129:127–137, 2015.

[84] Paul Rayson and Roger Garside. *"Comparing Corpora Using Frequency Profiling"*, volume 9, chapter Proceedings of the Workshop on Comparing Corpora. Association for Computational Linguistics, 2002.

[85] Majid Khosravinik e Michal Krzyzanowski e Tony McEnery e Ruth Wodak Paul Baker, Costas Gabrielatos. A useful methodological synergy? combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the uk press. *Discourse & Society*, 19:273–306, 2008.

[86] Lisa Kaati e Jonas Clausen Mork Katie Cohen, Fredrik Johansson. Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence*, 26:246–256, 2014.

[87] Anna Korhonen. Automatic lexical classification – balancing between machine learning and linguistics. In Olivia Kwong, editor, *23rd Pacific Asia Conference on Language, Information, and Computation*, page 19–28, 2009.

[88] Eliot Higgins. *"Geolocation Techniques—Mapping Landmarks"*. Bellingcat, 2014.

[89] Jukka Matthias Krisp e Liqiu Meng Andreas Hackeloeer, Klaas Klasing. Georeferencing: a review of methods and applications. *Annals of GIS*, 20:61–69, 2014.

[90] What is geospatial intelligence (geoint)? definition and faqs, 2022. URL `https://www.heavy.ai/technical-glossary/geoint`.

[91] Geospatial intelligence, 2022. URL `https://www.satcen.europa.eu/page/geospatial_intelligence`.

[92] Commissione di Studio GeoInt. Imint e geoint, qual è la differenza? ne parla la commissione geoint, Apr 2021. URL `https://news.socint.org/la-differenza-tra-imint-e-geoint/`.

[93] Quirine A.M. Eijkman and D. J. Weggemans. Open source intelligence and privacy dilemmas: Is it time to reassess state accountability? *Security and Human Rights*, 23:285–296, 2013.

[94] Jan-Jaap Oerlemans. Privacy risks of (automated) open source intelligence (osint), Febbraio 2022. URL `https://aboutintel.eu/privacy-and-automated-osint/`.

[95] Peride K. Blind. Accountability in public service delivery: A multidisciplinary review of the concept. In *Expert Group Meeting Engaging Citizens to Enhance Public Sector Accountability and Prevent Corruption in the Delivery of Public Services*, Vienna, 2011.

[96] Quirine A.M. Eijkman e B. T. van Ginkel. Compatible or incompatible? intelligence and human rights in terrorist trials. *Amsterdam Law Forum*, 2011.

[97] Wikileaks. Stratfor emails: Wikileaks impact is stratfor's bottom line, 2012. URL http://wikileaks.org/WikiLeaks-Impact-is-Stratfor-s.html.

[98] Charles Arthur. Trapwire surveillance system exposed in document leak. *The Guardian*, 2012. URL https://www.theguardian.com/world/2012/aug/13/trapwire-surveillance-system-exposed-leak.

[99] Ryan Gallagher. Software that tracks people on social media created by defense firm. *The Guardian*, 2013. URL https://www.theguardian.com/world/2013/feb/10/software-tracks-social-media-defence.

[100] Public Intelligence. Unravelling trapwire: The cia-connected global suspicious activity surveillance system, 2012. URL https://publicintelligence.net/unravelling-trapwire/.

[101] Ben Hayes. Spying in a see through world: The 'open source' intelligence industry, 2010. URL https://www.statewatch.org/statewatch-database/spying-on-a-see-through-world-the-open-source-intelligence-industry-by-ben-hayes/.

[102] John Markof. Entrepreneurs see a web guided by common sense. *New York Times*, November 12 2006.

[103] Matt Burgess. If russia invades ukraine, tiktok will see it up close. *Wired Magazine*, Febbraio 2022. URL https://www.wired.co.uk/article/russia-ukraine-military-photos-video.

[104] Chris Looft e Desiree Adib. The independent investigators tracking russia's military buildup. *ABC News*, Gennaio 2022. URL https://abcnews.go.com/Technology/independent-investigators-tracking-russias-military-buildup/story?id=82529068.

[105] Centre for Information Resilience, 2022. URL https://www.info-res.org/.

[106] bellingcat, 2022. URL https://www.bellingcat.com/about.

[107] Eliot Higgins, Febbraio 2022. URL https://twitter.com/EliotHiggins/status/1495355366141534208.

[108] Dpr prevents several blasts attempted by ukrainian saboteurs, Febbraio 2022. URL https://tass.com/emergencies/1405995.

[109] @oldLentach, Febbraio 2022. URL https://twitter.com/oldLentach/status/1494962375816007685.

[110] Russian border security eliminates five saboteurs infiltrating from ukraine, Febbraio 2022. URL https://tass.com/emergencies/1407169.

[111] Febbraio 2022. URL https://youtu.be/tyCK_hdWxdE.

[112] Firms: Fire information for resource management system, 2022. URL https://firms.modaps.eosdis.nasa.gov/map/#t:adv;d:2022-02-21;@38.3,47.3,12z.

[113] BBC. False flags: What are they and when have they been used?, Febbraio 2022. URL https://www.bbc.com/news/world-60434579.