IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SECCIÓN ESPECIAL SOBRE ENFOQUES EMERGENTES A LA SEGURIDAD CIBERNÉTICA

Recibido el 19 de diciembre de 2019, aceptado el 5 de enero de 2020, fecha de publicación 9 de enero de 2020, fecha de la versión actual 16 de enero de 2020.

Identificador de objeto digital 10.1109/ACCESS.2020.2965257

La mina de oro aún no explotada de OSINT: Oportunidades, Desafíos abiertos y tendencias futuras

JAVIER PASTOR GALINDO, PANTALONE NÉSPOLI, FÉLIX GÓMEZ MÁRMOL, Y GREGORIO MARTÍNEZ PÉREZ

Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, 30100 Murcia, España

Autor para correspondencia: Javier Pastor-Galindo (javierpg@um.es)

Este trabajo ha sido financiado en parte por el Contrato Predoctoral FPU (FPU18/00304) concedido por el Ministerio de Ciencia, Innovación y Universidades de España, en parte por el Contrato Predoctoral FPU concedido por la Universidad de Murcia, en parte por el Instituto de Investigación Ramón y Cajal Contrato (RYC-2015-18210) concedido por el MINECO (España) y cofinanciado por el Fondo Social Europeo, y en parte por el proyecto SAFEMAN *(Un marco de gestión unificado para la ciberseguridad y la seguridad en la industria manufacturera)* con código *RTI2018- 095855-B-I00.*

RESUMEN La cantidad de datos que genera el mundo interconectado actual es inconmensurable, y gran parte de esos datos están disponibles públicamente, lo que significa que son accesibles para cualquier usuario, en cualquier momento y desde cualquier lugar de Internet. En este sentido, Open Source Intelligence (OSINT) es un tipo de inteligencia que en realidad se beneficia de esa naturaleza abierta al recopilar, procesar y correlacionar puntos de todo el ciberespacio para generar conocimiento. De hecho, los avances tecnológicos recientes están provocando que OSINT evolucione actualmente a un ritmo vertiginoso, proporcionando aplicaciones innovadoras basadas en datos y potenciadas por IA para la política, la economía o la sociedad, pero también ofreciendo nuevas líneas de acción contra las ciberamenazas y el ciberdelito. El artículo en cuestión describe el estado actual de OSINT y hace una revisión exhaustiva del paradigma, centrándose en los servicios y técnicas que mejoran el campo de la ciberseguridad. Por un lado, analizamos los puntos fuertes de esta metodología y proponemos numerosas formas de aplicarla a la ciberseguridad. Por otro lado, cubrimos las limitaciones a la hora de adoptarlo. Teniendo en cuenta que queda mucho por explorar en este amplio campo, también enumeramos algunos desafíos abiertos para abordar en el futuro. Adicionalmente, estudiamos el rol de OSINT en la esfera pública de los gobiernos, que constituyen un panorama ideal para explotar los datos abiertos.

ÍNDICE TÉRMINOS OSINT, ciberinteligencia, ciberseguridad, ciberdefensa, desafíos, seguridad nacional, delitos informáticos, inteligencia computacional, adquisición de conocimiento, servicios de redes sociales, herramientas de software, privacidad de datos, Internet.

I. INTRODUCCIÓN

Open Source Intelligence (OSINT) consiste en la recopilación, procesamiento y correlación de información pública de fuentes de datos abiertas como medios de comunicación, redes sociales, foros y blogs, datos de gobiernos públicos, publicaciones o datos comerciales. Dados algunos datos de entrada, junto con la aplicación de técnicas avanzadas de recopilación y análisis, OSINT amplía continuamente el conocimiento sobre el objetivo. De esta manera, la información encontrada alimenta nuevamente el proceso de recolección para acercarse al objetivo final [1].

Hoy en día, OSINT es ampliamente adoptado por los gobiernos y los servicios de inteligencia para realizar sus investigaciones y luchar contra el ciberdelito [2]. Sin embargo, no solo se utiliza para asuntos de Estado, sino que se aplica a varios objetivos diferentes.

El editor asociado que coordina la revisión de este manuscrito y quien lo aprobó para su publicación fue Luis Javier García Villalba.

De hecho, la investigación actual se centra en (pero no se limita a) tres aplicaciones principales que se representan en la FIGURA 1 y se describen a continuación:

- *Opinión social y análisis de sentimientos:* junto con el auge de las redes sociales en línea, es posible recopilar las interacciones de los usuarios. , mensajes, intereses y preferencias para extraer conocimiento no explícito. La evidencia acumulada de las redes sociales es de gran alcance y muy ventajosa [3]. Dicha recopilación y análisis podría aplicarse, por ejemplo, al marketing, las campañas políticas o la gestión de desastres [4].

- *Ciberdelincuencia y delincuencia organizada:* los datos abiertos se analizan y comparan continuamente mediante procesos OSINT para detectar intenciones delictivas en una etapa temprana. Teniendo en cuenta los patrones de los adversarios y las relaciones entre delitos graves, OSINT puede brindar a las fuerzas de seguridad la oportunidad de detectar rápidamente

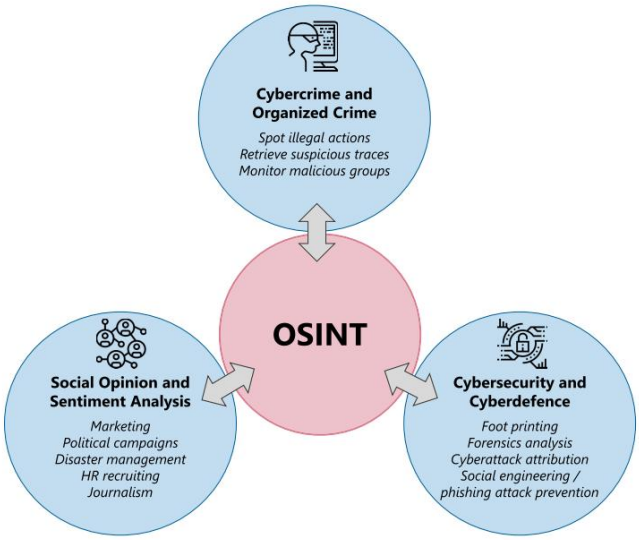

FIGURA 1. Principales casos de uso de OSINT.

acciones ilegales [5]. En esta dirección, mediante la explotación de los datos abiertos sería posible rastrear la actividad de las organizaciones terroristas, cada vez más activas en Internet [6], [7]. • *Ciberseguridad y ciberdefensa: Los sistemas* TIC *(Tecnologías de la Información y la Comunicación)* son atacados continuamente por delincuentes con el objetivo de interrumpir la disponibilidad de los servicios prestados [8]. La investigación se vuelve, por lo tanto, crucial para defender esos sistemas de los ciberatacantes, concretamente para hacer frente a los desafíos que aún están abiertos en el campo de la ciberseguridad [9]. En este sentido, las ciencias de datos no solo se están aplicando para el footing en pentestings, sino también para la protección preventiva de organizaciones y empresas. Concretamente, las técnicas de minería de datos pueden ayudar a realizar análisis de los ataques diarios, correlacionarlos y respaldar los procesos de toma de decisiones para una defensa eficaz, pero también para una reacción rápida [10]. Del mismo modo, OSINT también puede considerarse en este contexto como una fuente de información para rastreos e investigaciones. El análisis digital forense [11] puede incorporar OSINT para complementar las evidencias digitales dejadas por un incidente.

Además de estos, OSINT se puede aplicar a otros contextos. En particular, se puede extraer información relevante realizando ataques de ingeniería social. Las entidades mal motivadas aprovechan la información disponible públicamente publicada en línea (por ejemplo, en las redes sociales) para crear ganchos atractivos para capturar al objetivo [12]. Además, es posible realizar una evaluación automática de la veracidad de los datos abiertos con el objetivo de revelar noticias falsas y falsificaciones profundas, entre otros [13].

No obstante, es importante notar que la utilización de datos públicos también tiene problemas comprometedores. Por un lado, el Reglamento General de Protección de Datos de la UE (RGPD) limita el procesamiento de datos personales relacionados con personas en la zona de la UE [14]. Por otro lado, existe un fuerte componente ético que está ligado a la privacidad de los usuarios. En particular, la elaboración de perfiles de personas [15] podría revelar datos personales como su preferencia política, orientación sexual o creencias religiosas, entre otros. Además, la explotación de una cantidad tan grande de información puede dar lugar a abusos, lo que resultaría en daño a inocentes a través del acoso cibernético, los chismes cibernéticos o las agresiones cibernéticas [16].

El trabajo que nos ocupa, que es una extensión del trabajo propuesto en [17], abarca el presente y el futuro de OSINT analizando sus puntos positivos y negativos, describiendo formas de aplicar este tipo de inteligencia y enunciando direcciones futuras para la evolución. de este paradigma. Además, en este trabajo se presenta una descripción más detallada de diferentes técnicas, herramientas y desafíos abiertos. Además, proponemos la integración de OSINT dentro del modelo DML *(Detection Maturity Level)* para abordar el problema de atribución desde una perspectiva diferente en el contexto de las investigaciones de ciberataques. También presentamos flujos de trabajo de muestra para facilitar la comprensión y el uso de OSINT para recopilar información valiosa a partir de entradas básicas.

Además, nuestro propósito es estimular las investigaciones y avances en el ecosistema OSINT. El alcance de dicho ecosistema es bastante amplio, abarcando desde la psicología, las ciencias sociales hasta la contrainteligencia y el marketing. Como hemos visto hasta ahora, OSINT es un mecanismo prometedor que mejora concretamente los campos tradicionales de ciberinteligencia, ciberdefensa y forense digital [18]. El impacto que esta metodología podría tener en la sociedad gracias a la tecnología actual y la gran cantidad de fuentes abiertas aún está sin explotar. Todavía queda un largo camino por explorar en este tema, y este artículo presenta algunas líneas de investigación atractivas para el futuro.

El resto de este documento está organizado de la siguiente manera. La SECCIÓN II ofrece una revisión de trabajos de investigación recientes en el campo de OSINT. La SECCIÓN III analiza la motivación, los pros y los contras del desarrollo de OSINT. LA SECCIÓN IV explica los principales pasos de OSINT y los flujos de trabajo prácticos para llevarlos a cabo. Luego, la SECCIÓN V incluye una descripción detallada de las técnicas y servicios de recopilación basados en OSINT. La SECCIÓN VI analiza y compara algunas herramientas OSINT que automatizan la recopilación y el análisis de información OSINT. LA SECCIÓN VII propone la integración de OSINT en la investigación de ciberataques. La SECCIÓN VIII se enfoca en el impacto de OSINT dentro de una nación, no solo en aras de sus operaciones internas de ciberdefensa, sino también como beneficiario de políticas de transparencia. España se toma específicamente como referencia de afinidad y se contextualiza con el resto del mundo. LA SECCIÓN IX plantea algunos desafíos abiertos en relación con la investigación en OSINT. Finalmente, la SECCIÓN X concluye con algunos comentarios clave, así como futuras direcciones de investigación.

II. ESTADO DEL ARTE

En los últimos años, con los avances de las técnicas de big data y minería de datos, la comunidad investigadora ha notado que los datos abiertos representan una fuente poderosa para analizar comportamientos sociales y obtener información relevante [19]. A continuación, describimos algunos trabajos notables que giran en torno a cada uno de los tres casos de uso principales mencionados anteriormente para OSINT.

Con respecto al uso de OSINT para **extraer opinión social y emociones,** Santarcangelo *et al.* [20] propusieron un modelo para determinar las opiniones de los usuarios sobre una determinada palabra clave a través de las redes sociales, estudiando específicamente los adjetivos, intensificadores y negaciones utilizados en los tuits. Desafortunadamente, es una solución simple basada en palabras clave diseñada solo para el idioma italiano, sin tener en cuenta los problemas semánticos. Por otro lado, Kandias *et al.* [21] podría relacionar el uso de las personas de las redes sociales (en particular, Facebook) con su nivel de estrés. Sin embargo, los experimentos se llevaron a cabo solo con 405 usuarios, mientras que hoy en día existe la posibilidad de procesar cantidades de datos mucho mayores. Otro estudio interesante se lleva a cabo en [22], donde los autores aplicaron el procesamiento del lenguaje natural (NLP) a los mensajes de WhatsApp para posiblemente prevenir la ocurrencia de violencia masiva en Sudáfrica.

Desafortunadamente, la investigación se limita a mensajes de texto, por lo que se excluye información vital que puede divulgarse a través de material multimedia.

En el contexto del **cibercrimen y el crimen organizado,** existen varios trabajos que exploran la aplicación de OSINT para investigaciones criminales [23]. Por ejemplo, OSINT podría aumentar la precisión de los procesamientos y arrestos de los culpables con marcos como el propuesto por Quick y Choo [11]. Concretamente, los autores aplican OSINT a datos forenses digitales de una variedad de dispositivos para mejorar el análisis de inteligencia criminal. En este campo, otra oportunidad que brinda OSINT es la detección de acciones ilícitas así como la prevención de futuros delitos como atentados terroristas, asesinatos o violaciones. De hecho, los proyectos europeos ePOOLICE [24] y CAPER [25] fueron diseñados para desarrollar modelos efectivos para escanear datos abiertos automáticamente con el fin de analizar la sociedad y detectar el crimen organizado emergente. A diferencia de los proyectos anteriores, cuyas propuestas no fueron prácticamente utilizadas en casos reales, Delavallade *et al.* [26] describen un modelo basado en datos de redes sociales que es capaz de extraer indicadores de delincuencia futuros. Dicho modelo se aplica luego al robo de cobre ya los casos de uso de propaganda yihadista.

Desde el punto de vista de la **ciberseguridad y la ciberdefensa,** OSINT representa una valiosa herramienta para mejorar nuestros mecanismos de protección frente a los ciberataques.
Hernández *et al.* [27] proponen el uso de OSINT en el contexto colombiano para prevenir ataques y permitir la anticipación estratégica. Incluye no solo complementos para recopilar información, sino también modelos de aprendizaje automático para realizar análisis de opinión. Además, el proyecto europeo DiSIEM [28] mantiene como primer objetivo la integración de diversas fuentes de datos OSINT en los actuales sistemas SIEM *(Security Information and Event Management)* para ayudar a reaccionar ante vulnerabilidades recientemente descubiertas en la infraestructura o incluso predecir posibles amenazas emergentes. . Además, Lee y Shon [29] también diseñaron un marco basado en OSINT para inspeccionar las amenazas de ciberseguridad de las redes de infraestructura crítica. Sin embargo, todos estos enfoques no se han aplicado a escenarios del mundo real, por lo que su eficacia sigue siendo cuestionable.

Extendiendo la disertación a otros campos de aplicación, en [30] los autores demuestran cómo recordar pasivamente información significativa sobre los empleados de la organización de forma automatizada. Dicha información se relaciona luego con el análisis de la denominada *superficie de ataque de ingeniería social,* mostrando la factibilidad efectiva del enfoque propuesto. Luego, los autores proponen un conjunto de contramedidas potenciales, incluido un escáner de vulnerabilidades de ingeniería social disponible públicamente que las empresas pueden aprovechar para reducir la exposición de sus empleados.

Además, en [31] se realiza una revisión sistemática de los enfoques, metodologías y herramientas propuestas por la academia para realizar una evaluación automática de la veracidad de los datos disponibles públicamente. En concreto, los autores estudiaron 107 artículos de investigación entre 2013 y 2017 para argumentar sobre el estado del arte de la evaluación de la veracidad, que se ha convertido en una gran preocupación durante la última década debido a la difusión de fake news y deepfakes. En esta dirección, los autores destacan la relativa inmadurez de este campo, identificando varios desafíos que caracterizarán las futuras tendencias de investigación.

tercero VENTAJAS Y DEFICIENCIAS DE OSINT

Los campos de aplicación de OSINT son numerosos y las soluciones que se desarrollan bajo este paradigma son cada vez mayores.
Sin embargo, detrás de esta metodología hay una compensación con la que los desarrolladores e ingenieros tienen que lidiar. Desde un punto de vista técnico, como podemos ver en la TABLA 1, OSINT expone una serie de beneficios, pero también tiene que lidiar con algunas restricciones, que se detallan a continuación.

A. BENEFICIOS DE OSINT

1) ENORME CANTIDAD DE INFORMACIÓN DISPONIBLE

Actualmente hay un gran volumen de datos de fuente abierta que vale la pena analizar, correlacionar y vincular [32]. Esto incluye redes sociales, documentos e informes de gobiernos públicos, contenido multimedia en línea, periódicos e incluso la Deep web y la Dark web [33], entre otros. De hecho, tanto la Deep Web como la Dark Web (esta última circunscrita dentro de la primera) contienen incluso más información que la Surface Web (es decir, la Internet conocida por la mayoría de los usuarios) [34]. Para poder acceder a estas redes es necesario utilizar herramientas específicas ya que sus contenidos no están indexados por los buscadores tradicionales.

A diferencia de Surface Web y la mayoría de Deep Web, Dark Web ofrece anonimato y privacidad a los usuarios que la utilizan. Esta propiedad facilita que los delincuentes utilicen esta red para navegar, realizar sus búsquedas y publicar con fines ilegítimos mientras ocultan su identidad. Por tanto, la Dark Web es una fuente ideal para aplicar OSINT y luchar contra el cibercrimen, el crimen organizado o las ciberamenazas. Por otro lado, la búsqueda y desanonimización de estas personas son desafíos actuales no triviales para que OSINT funcione correctamente [35].

2) ALTA CAPACIDAD DE CÓMPUTO

Los avances en la arquitectura informática, los procesadores y las GPU (unidades de procesamiento gráfico) permiten realizar trabajos intensivos en mano de obra

TABLA 1. Pros y contras de OSINT en pocas palabras.

Pros ✓	Cons ✗
Huge amount of available information	Complexity of data management
High capacity of computing	Unstructured information
Big data and machine learning	Misinformation
Complementary types of data	Data sources reliability
Flexible purpose and wide scope	Strong ethical/legal considerations

operaciones en términos de recolección, procesamiento, análisis y almacenamiento [36]. Gracias a esta característica, tenemos la oportunidad de aplicar OSINT considerando grandes cantidades de información pública y mezclando una gran cantidad de conjuntos de datos, relaciones y patrones de diferentes tipos de fuentes abiertas, mientras aplicamos técnicas avanzadas de procesamiento y análisis.

3) BIG DATA Y MACHINE LEARNING

Proliferación emergente de técnicas de análisis y minería de datos, así como algoritmos de aprendizaje automático, que pueden automatizar y hacer más inteligentes y eficientes los procesos de investigación y toma de decisiones [36]. Permite detectar correlaciones complejas que son naturalmente impredecibles para los humanos. Este punto será clave en las futuras actividades de OSINT, ya que marcará la diferencia entre la investigación dirigida por humanos y la dirigida por inteligencia artificial. Al incorporar estas técnicas, definitivamente mejorará el proceso de recolección y análisis, lo que resultará en investigaciones precisas y cercanas a nuestro objetivo. Además, las agencias gubernamentales de contrainteligencia pueden aprovechar dicho paradigma para mejorar aún más la calidad de la información administrada y, en consecuencia, la batalla contra las organizaciones terroristas [37].

4) TIPOS DE DATOS COMPLEMENTARIOS

Posibilidad de alimentar OSINT con otros tipos de información [38]. La estructura inherente del sistema es lo suficientemente abierta como para incluir datos que en realidad no se han obtenido de fuentes abiertas. Este hecho significa que OSINT puede ser aún más efectivo si somos capaces de agregar información externa para complementar las investigaciones. Por ejemplo, las agencias de aplicación de la ley podrían aprovechar la colaboración de los ciudadanos para alimentar las búsquedas de OSINT, los servicios de inteligencia podrían aprovechar información clasificada sobre ciberdelincuentes o incidentes para enriquecer las investigaciones de OSINT, o incluso los usuarios comunes podrían combinar OSINT con ingeniería social para perfilar su objetivo.

5) PROPÓSITO FLEXIBLE Y AMPLIO ALCANCE

Debido a la naturaleza de OSINT, las investigaciones pueden extenderse a muchos problemas y pueden recopilar información en todo el ciberespacio. Este paradigma podría ser utilizado para aspectos económicos, psicológicos, estratégicos, periodísticos, laborales o de seguridad, entre otros. En particular, podríamos destacar los beneficios en el campo del crimen y la ciberseguridad, donde OSINT podría monitorear personas sospechosas o grupos peligrosos, detectar perfiles de influencia relacionados con la radicalización, estudiar preocupantes tendencias de la sociedad, apoyar la atribución de ciberataques y delitos, potenciar el análisis forense digital, etc. [5], [18].

B. LIMITACIONES DE OSINT

1) COMPLEJIDAD DE LA GESTIÓN DE DATOS

La cantidad de datos es enorme y, en consecuencia, es un desafío manejarlos de manera eficiente y efectiva [39]. Es beneficioso para OSINT considerar la mayor cantidad de información posible, pero también contar con técnicas avanzadas y recursos significativos para garantizar una recopilación, procesamiento y análisis de alta calidad.

2) INFORMACIÓN NO ESTRUCTURADA

La información pública disponible en Internet está inherentemente desorganizada masivamente. Esto significa que los datos recopilados por OSINT son tan heterogéneos que dificulta clasificarlos, vincularlos y examinarlos para extraer relaciones y conocimientos relevantes [4]. En este sentido, OSINT requiere de mecanismos como la minería de datos, el Procesamiento del Lenguaje Natural (NLP) o la analítica de texto para homogeneizar la información no estructurada con el fin de poder explotarla.

3) DESINFORMACIÓN

Las redes sociales y los medios de comunicación se inundan de opiniones subjetivas, *fake news* y patrañas [4]. Por esta razón, la existencia de información inexacta debe tenerse en cuenta en la implementación de los mecanismos OSINT y no debe impulsar la propagación de la búsqueda. Las actividades de OSINT siempre deben tratar con información confiable y seguir líneas de exploración confiables para garantizar resultados positivos y convincentes [40].

4) CONFIABILIDAD DE LAS FUENTES DE DATOS

La confiabilidad y la autoridad de la información son, de hecho, la clave para el éxito de las investigaciones OSINT [41]. Idealmente, los datos recopilados deben provenir de fuentes autorizadas, revisadas y confiables (documentos oficiales, informes científicos, medios de comunicación confiables) [39]. En la práctica, OSINT coexistirá también con fuentes subjetivas o no autorizadas, como el contenido de las redes sociales o los medios manipulados [42]. Si bien este tipo de fuentes es más propensa a la desinformación, en realidad es de donde más conocimiento se puede extraer para investigar a personas, grupos o empresas. Si la credibilidad de las fuentes abiertas de información representa una limitación, se vuelve aún más desafiante considerando la posible ambigüedad de las consultas de los usuarios para recuperar la información deseada [43].

5) FUERTES CONSIDERACIONES ÉTICAS/LEGALES

Numerosas preocupaciones sobre la privacidad, el respeto y la integridad personal surgen con el desarrollo de OSINT [44]. En este sentido, cabe señalar que la cuestión de si OSINT constituye una cuestión ética se sitúa generalmente en el ámbito de la ética de la recopilación de inteligencia [45].

Por un lado, aunque es de acceso público, OSINT tiene el poder de divulgar información que no se publica explícitamente en la web. Los resultados descubiertos deben respetar la privacidad de los usuarios y no revelar cuestiones íntimas y personales [15], teniendo en cuenta las reglamentaciones relacionadas actuales (como GPDR [14]).
En este sentido, de Internet se pueden inferir aspectos como la orientación sexual, las creencias religiosas, la inclinación política o las conductas comprometedoras, y este proceso de divulgación puede resultar problemático en muchos países en la actualidad. Por otro lado, el alcance de las búsquedas basadas en OSINT debería estar, por definición, limitado a fuentes de datos abiertas. Bajo ninguna circunstancia se pueden eludir los controles de acceso o los métodos de autenticación para extraer conocimiento.

IV. FLUJOS DE TRABAJO

OSINT OSINT, como cualquier otro tipo de inteligencia, tiene una metodología bien definida y precisa. Desde nuestro punto de vista científico-técnico, nos interesan especialmente tres pasos.

En primer lugar, en la fase de **recopilación ,** los datos disponibles públicamente se recuperan de fuentes abiertas relevantes según el objetivo u objetivo. En particular, Internet es el recurso por excelencia por el volumen de material existente y su fácil accesibilidad. El proceso de recolección es particularmente relevante porque a partir de esta etapa se desencadena todo el proceso de generación de inteligencia.

Luego, en la fase de **análisis ,** la materia prima recolectada es tratada para generar información valiosa y comprensible.
Los datos por sí solos no son útiles, por lo que hay que interpretarlos para obtener los primeros hechos derivados de un análisis en profundidad.

Finalmente, en el proceso de **extracción de conocimiento ,** la información previamente depurada se toma como entrada para algoritmos de inferencia más sofisticados. Gracias a los avances computacionales de la era actual, es posible detectar patrones, perfilar comportamientos, predecir valores o correlacionar eventos.

Cabe mencionar que el segundo y tercer paso comprenden tecnologías ampliamente utilizadas y conocidas en el contexto de la minería de datos. Sin embargo, el enfoque de recopilación OSINT difiere de los servicios basados en datos actuales. Hoy en día, las aplicaciones comunes de análisis de datos recopilan la mayor cantidad de información posible de fuentes de datos predefinidas e implementan procesos de recopilación claros. Por el contrario, las soluciones OSINT deben recopilar datos específicos del mar de todos los posibles y accesibles abiertos. recursos.

Para enfrentar esta última incertidumbre desafiante e ir un paso más allá, en la FIGURA 2 proponemos un marco práctico para llevar a cabo investigaciones basadas en OSINT. Hemos incluido aquellas rutas de exploración que vale la pena seguir para optimizar el análisis de los resultados de la colección y maximizar la extracción de conocimiento. Esta alta abstracción esquema incluye las transacciones más claras, los elementos representativos y las operaciones pendientes.

A. COLECCIÓN OSINT

Antes de los pasos de análisis y extracción de inteligencia, el investigador debe ampliar el conjunto de datos sobre el objetivo.
Con este objetivo, proponemos algunas técnicas OSINT para representar diferentes estrategias de recolección. En particular, hemos considerado las técnicas OSINT *de los motores de búsqueda, las redes sociales, la dirección de correo electrónico, el nombre de usuario, el nombre real, la ubicación, la dirección IP* y *el nombre de dominio* (como describiremos más detalladamente en la SECCIÓN V). Debajo de cada uno, habrá innumerables servicios OSINT con formas similares de recopilar datos.

En esta fase, se supone que, al menos, está disponible un dato atómico sobre el objetivo (por ejemplo, nombre real, nombre de usuario, dirección de correo electrónico, etc.). A partir de esa semilla inicial y según su naturaleza, el investigador aplica las técnicas OSINT más adecuadas para derivar más datos. En este sentido, los resultados obtenidos con una determinada técnica son una *transferencia de datos* para ser utilizados por otro tipo de técnica. Estas transacciones representadas ilustran posibles formas de propagar la investigación, donde la salida de la técnica de origen se convierte en la entrada para alimentar la técnica de destino.

B. ANÁLISIS OSINT

Las iteraciones continuas a través de las diferentes técnicas OSINT deben analizarse y comprenderse para generar información valiosa. Hay una cantidad creciente de técnicas de análisis en la literatura para realizar esta tarea [46], destacando a continuación aquellos procedimientos atractivos que son aplicables en nuestro escenario: • *Análisis léxico:* los datos sin procesar deben examinarse para extraer entidades y relaciones del texto.
Es fundamental aplicar procesos de traducción al lenguaje utilizado en la investigación OSINT [47] y filtrar el ruido que no agrega valor de las oraciones que no agregan valor. • *Análisis semántico:* De nada sirve tener una bolsa de palabras si no se extrae el significado [48]. Con este fin de comprender los datos, hoy en día se utilizan algoritmos de procesamiento de lenguaje natural [49]. Además, las técnicas de análisis de sentimientos permiten contextualizar publicaciones u opiniones subjetivas para clasificar el estado emocional del autor (p. ej., positivo, negativo o neutral).

Finalmente, los procedimientos de descubrimiento de la verdad abordan la desafiante tarea de resolver conflictos en datos de múltiples fuentes que se encuentran en posiciones opuestas sobre el mismo tema [50]. • *Análisis geoespacial:* Vale la pena analizar los datos recopilados de redes sociales, eventos, sensores o direcciones IP desde una perspectiva basada en la ubicación. En este sentido, el uso de mapas o gráficos facilita la representación y comprensión de los datos [51], así como la extracción de conexiones significativas entre incidentes o personas

• *Análisis de redes sociales:* las características que aportan las redes sociales modernas permiten a los investigadores realizar un análisis en profundidad de los usuarios [52]. En tal escenario, el análisis de

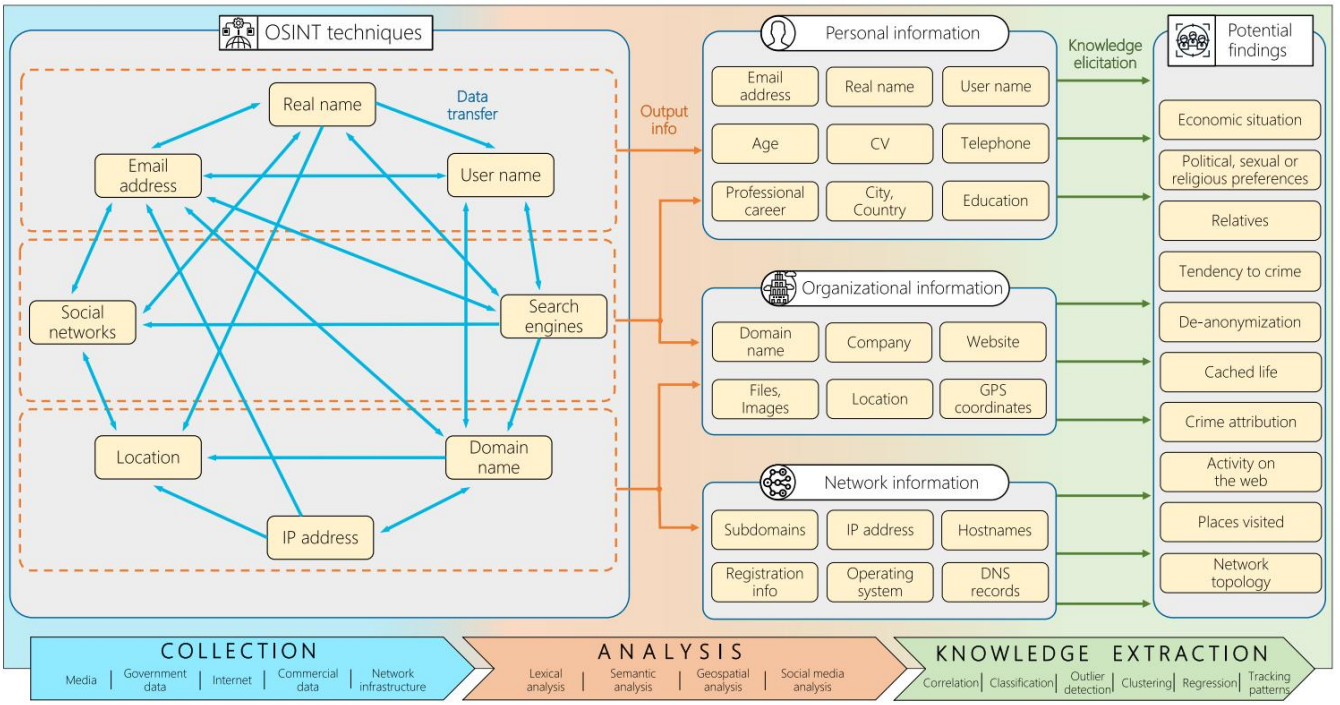

FIGURA 2. Principales flujos de trabajo OSINT e inteligencia derivada.

Los datos sociales permiten la creación de una red de contactos, interacciones, lugares, comportamientos y gustos en torno al tema.

Los resultados de la puesta en marcha de las técnicas antes mencionadas se consideran *información de salida* y se clasifican en tres grupos principales:

- La *información personal* fusiona los detalles de identidad de la persona que se obtienen principalmente del nombre real, la dirección de correo electrónico, el nombre de usuario, las redes sociales y las técnicas de los motores de búsqueda. .

- La *información organizacional* está formada por aspectos de un equipo o empresa compuesta por individuos. Se recopila esencialmente a través de técnicas de redes sociales, motores de búsqueda, ubicación, nombre de dominio y dirección IP.

- La *información de la red* cubre los datos técnicos de los sistemas y las topologías de comunicación que generalmente se logran a través de técnicas de ubicación, nombre de dominio y dirección IP.

Lógicamente, estos tres bloques de información se pueden ampliar con más elementos. Además, una sola investigación puede tener diferentes tipos de *información de salida* que se complementan entre sí.

C. EXTRACCIÓN DE CONOCIMIENTO OSINT

El valor de la información recopilada hasta el momento es incuestionable. Sin embargo, la extracción de inteligencia de esos hallazgos conduce en realidad a lo que proporcionará un reconocimiento atractivo del objetivo [53]. Para ello, consideramos la *elicitación de conocimiento* como el tratamiento de los resultados del análisis *(output info)* haciendo uso de la minería de datos y artificial.

técnicas de inteligencia. A continuación mencionamos algunas tecnologías realmente prometedoras en esta etapa:

- *Correlación:* Detección de relaciones entre personas, eventos o datos en general [54]. Las características fuertemente relacionadas son especialmente valiosas para revelar aquellas asociaciones no explícitas que existen en el conjunto de datos.
- *Clasificación:* Los datos se pueden dividir en grupos según categorías predefinidas (aprendizaje supervisado) [55]. Esta técnica permite la organización de grandes cantidades de información para una extracción de conocimiento más eficaz [56].

- *Detección de valores atípicos:* este procedimiento analiza el conjunto de datos y detecta anomalías en él [57]. Son particularmente interesantes para la observación de agentes malignos, cuyo comportamiento o acciones difieren de la población general.
- *Clustering:* Asigna piezas de datos en clusters, pudiendo considerar gran cantidad de condiciones o heurísticas [58]. Esto podría revelar, por ejemplo, diferentes formas de comportarse en la red, varios tipos de perfiles en línea o categorizar formas de atacar a personas, organizaciones o infraestructuras [59] sin conocer de antemano la existencia de esa diversidad (aprendizaje no supervisado).

- *Regresión:* El principal objetivo de esta técnica es pronosticar o predecir valores numéricos o hechos [60]. Por ejemplo, una regresión lineal devuelve un valor atendiendo a una función lineal, una red neuronal es una estructura que mapea combinaciones complejas de entradas a una salida, o aprendizaje profundo que se compone de varias capas que se combinan y realizan operaciones con la entrada.

• *Seguimiento de patrones:* a diferencia de la detección de anomalías, el reconocimiento de patrones es un proceso para detectar regularidades en los datos [61]. Los métodos mencionados anteriormente pueden incluirse en este concepto amplio de descubrimiento de conocimiento. De hecho, cualquier técnica de inteligencia artificial es apta para la extracción de conocimiento con datos abiertos.

Estas técnicas inteligentes permiten inferir cuestiones abstractas, complejas y jugosas sobre el target que no están explícitamente publicadas en Internet [62]. Sin embargo, este proceso plantea varios desafíos, que residen principalmente en investigar y desarrollar este proceso de extracción de conocimiento para identificar, perfilar o monitorear delincuentes, reconocer y explorar organizaciones maliciosas o descubrir y atribuir incidentes cibernéticos. Además, surgen varias consideraciones de privacidad debido a las poderosas inferencias que son potencialmente alcanzables.

El conocimiento extraído sobre una persona, empresa u organización puede ser especialmente sensible y su manipulación conduce indirectamente a problemas éticos y legales (específicamente tratados en el INCISO IX-F). De hecho, nunca debemos perder de vista el hecho de que estas técnicas podrían ser incluso mal utilizadas para dañar directamente a personas o grupos (análisis más profundo en la SUBSECCIÓN IX-G).

V. TÉCNICAS Y SERVICIOS DE COLECCIÓN DE OSINT

Como se ha demostrado, OSINT es bastante prometedor y poderoso, pero su implementación también es un desafío. . De hecho, la primera consideración es que precisa datos como punto de partida. Afortunadamente, el volumen de datos sin procesar no es un problema hoy en día debido a la existencia de Internet. Además, también hay un número creciente de aplicaciones, conocidas en este contexto como servicios OSINT, que precisamente facilitan la recopilación en la web.

A continuación, se presenta un resumen de las *técnicas OSINT más comunes.* Dentro de cada técnica, se muestran los *servicios OSINT* asociados más destacados en el momento de la redacción, dando pistas sobre cómo explotar eficazmente sus potencialidades. Cabe mencionar que los servicios OSINT son efímeros e incluso pueden aumentar o disminuir. Por el contrario, la técnica OSINT es un concepto más amplio que perdurará en el tiempo.

A. MOTORES DE BÚSQUEDA

Los buscadores *Google, Bing* o *Yahoo ,* entre otros, son herramientas muy conocidas y utilizadas. El uso tradicional de los mismos es la forma más sencilla de aplicar OSINT. Estos motores buscan dentro de la World Wide Web dada una consulta textual tratando de proporcionar información que coincida con la entrada, funcionando muy bien y devolviendo información valiosa al usuario.

Sin embargo, la cantidad de resultados puede ser tan abrumadora que incluso puede ser contraproducente para el usuario. Por esa razón, un buen investigador debe saber especificar las solicitudes dentro de un motor de búsqueda de acuerdo con el resultado deseado. Servicios como *Google* o *Bing* admiten filtros para refinar las búsquedas1 y recuperar exactamente el tipo de información que están interesados. Por ejemplo, el uso de "" permite coincidencias exactas, *OR* y *AND* actúan como operadores lógicos o comodines. También permite la introducción de condiciones como *filetype* para especificar un determinado tipo de archivo, *site* para limitar los resultados a los de un sitio web específico o *intitle* para encontrar páginas con ciertas palabras clave dentro de su título. La TABLA 2 contiene algunos operadores que se pueden usar para refinar las búsquedas de *Google* y *Bing* .

Yahoo, por su parte, no permite filtros específicos, pero podemos restringir la fecha, idioma o país de los resultados. El caso del buscador *DuckDuckGo* es especialmente interesante porque no rastrea al usuario, ni apunta a la dirección IP o al historial de búsqueda. Este enfoque de preservación de la privacidad hace que los resultados sean homogéneos para todos los usuarios, independientemente de sus hábitos, preferencias, ubicación o historial de búsqueda.

Además, algunos motores de búsqueda han sido diseñados para territorios específicos. *Yandex* es muy conocido en Rusia y Europa del Este e implementa operadores de búsqueda2 para restringir la búsqueda por URL, tipo de archivo, idioma, fecha, etc.
Baidu es otro servicio de búsqueda específico muy utilizado en Asia. Incluye no solo la típica barra de búsqueda de palabras clave, sino recursos adicionales dignos de OSINT como una red social, una sección de preguntas y respuestas, una biblioteca virtual o una enciclopedia, entre otros. También existen buscadores para la comunidad árabe como *Yamli* o *Eiktub,* pero están mucho menos empleados. Este tipo de servicios es particularmente interesante en investigaciones sobre personas, grupos y empresas pertenecientes a comunidades específicas.

Finalmente, es obligatorio conocer motores de búsqueda específicos para navegar en la Dark Web. Las investigaciones de OSINT contra el tráfico de drogas, la pornografía infantil, la venta de armas o el terrorismo se ven muy beneficiadas al explorar estos recursos no tan populares.
Con este fin, *Ahmia* y *Torch* son motores de búsqueda disponibles para su uso dentro de la red anónima Tor [63]. Sin embargo, el investigador tendrá que lidiar con el anonimato de esta red y sitios.

B. REDES SOCIALES

En la actualidad, es evidente la exposición del día a día de las personas y organizaciones en las redes sociales. Cualquier curioso se habrá dado cuenta de que se puede encontrar mucha información personal sin necesidad de conocimientos avanzados sobre estas plataformas.
Como se muestra en la TABLA 3, estas aplicaciones ofrecen posibilidades de búsqueda precisas en el contexto de OSINT. A continuación describimos algunas de las redes sociales más conocidas y utilizadas a nivel mundial.

Facebook es una red social repartida por todo el mundo con millones de usuarios. Podría considerarse un diario de sociedad, donde se puede encontrar información personal muy valiosa para las investigaciones de OSINT. El perfil de nuestro target puede revelar su empleo, educación, edad, ubicación, lugares visitados o grupos de agrado, entre otros. Las fotos y publicaciones también pueden ayudarnos a contextualizar la empresa o persona que estamos investigando, las zonas que frecuenta o el tipo de actividades que realiza. Además, también es posible buscar por

1https://support.google.com/websearch/answer/2466433

2https://yandex.com/support/search/query-language/operadores de búsqueda.html

TABLA 2. Algunos filtros de Google/Bing para búsqueda avanzada.

Google/Bing filter	Search operator	Example of use
Force an exact-match search	" "	"University of Murcia"
Exclude a term or phrase	-	university murcia -catholic
Search for X or Y	OR, |	university murcia|cartagena
Search for X and Y (used by default)	AND	university AND of AND murcia
Use of a wildcard	*	university of *
Search for a range of numbers	..	university murcia 2010..2019
Group terms or search operators	()	"university of (murcia|cartagena)"
Search within a given domain	site:	university murcia site:um.es
Search for a certain file type	filetype:	university murcia filetype:pdf
Search in page titles	intitle:	university intitle:umu
Search in URLs	inurl:	university inurl:um
Search in the text of the pages	intext:	university intext:murcia
Search the most recent cached version of a page	cache:	cache:um.es

TABLA 3. Potencialidades de diversas redes sociales.

Social Network	Type	Scope	Main potential for OSINT
4chan	Online community	Worldwide	Users interested in illicit activities
Badoo	Dating	Worldwide	Intimate and personal details
Cloob	Social connections	Iran	Personal profile, posting and community membership
Draugiem	Social connections	Latvia	Personal profile, publications in blogs, group membership
Facebook	Social connections	Worldwide	Personal profile, preferences and places visited
Facenama	Social connections	Iran	Personal profile, publications, photos and videos
Flickr	Photo-sharing	Worldwide	Activities, hobbies, places and personal relationships
Instagram	Social connections	Worldwide	Habits, locations and personal relationships
LinkedIn	Business	Worldwide	Professional profile, education, skills and languages
Mixi	Social connections	Japan	Personal profile, interests and opinions
Odnoklassniki	Social connections	Mainly Russia	Personal profile of adults, past and present friendships
Qzone	Social connections	Mainly China	Personal profile, preferences, habits
Reddit	Online community	Worldwide	Users trends, behaviors, and publications
Renren	Social connections	Mainly China	Personal profile of students, friendships and discussions
Taringa!	Social connections	Mainly Latin America	Personal profile, publications and community membership
Tinder	Dating	Worldwide	Intimate and personal details
Tumblr	Photo-sharing	Worldwide	Activities, hobbies, places and personal relationships
Twitter	Social connections	Worldwide	Personal profile, opinions and publications
VKontakte (VK)	Social connections	Mainly Russia	Personal profile, preferences and publications
Weibo	Social connections	Mainly China	Personal profile, opinions and publications
YouTube	Video-sharing	Worldwide	Video content, opinions and comments of subscribers

ubicación cuando no se conoce el nombre real, pudiendo finalmente encontrar el perfil de nuestro target.

YouTube es una plataforma basada en videos donde las grandes comunidades se conforman en torno a intereses compartidos. No solo es valioso el contenido subido por un usuario específico (temas, imágenes, escenas, lugares y personas que aparecen en los videos), sino también las opiniones y comentarios de los suscriptores.

Twitter se utiliza principalmente para la comunicación en vivo donde es común encontrar publicaciones personales a través de una línea de tiempo ordenada. Aparte de la información personal que revela el perfil, es especialmente interesante la extracción de las opiniones de los *tuits publicados,* las relaciones con los usuarios seguidos y seguidores o los *likes* en determinadas publicaciones. A partir de este tipo de interacciones, un investigador OSINT puede inferir la orientación del objetivo sobre ciertos temas, los intereses y preferencias de una organización, o cuán peligrosa puede ser una persona. Además, está disponible una interfaz fácil de usar[3] donde es posible buscar en toda la plataforma por palabras clave, frases exactas, hashtags, idioma, fecha, etc. Así, incluso podemos definir exploraciones a través de usuarios, menciones o respuestas.

Instagram también está muy extendido en la sociedad moderna como medio para compartir fotos. Los lugares, las personas y las actividades que se muestran en las imágenes también pueden ayudarnos a perfilar nuestro objetivo. La ubicación es un dato bastante sensible que se comparte con frecuencia en esta plataforma. En este sentido, también podemos

3 twitter.com/search-advanced

mencione servicios más específicos para compartir fotos como *Tumblr* o *Flickr.*

LinkedIn es el sitio más popular en el contexto de las redes sociales relacionadas con los negocios. Permite buscar por nombre real, empresa, organización, cargo o ubicación. En este caso, los perfiles profesionales pueden revelar datos de contacto completos, incluidas direcciones de correo electrónico y números de teléfono celular. Además, también podemos extraer información sobre el empleo, la educación, las habilidades, los idiomas y las relaciones comerciales.

También vale la pena considerar los sitios web de citas que se utilizan para contactar a personas en busca de pareja. A diferencia de otras redes sociales, donde muchos usuarios restringen sus datos personales, aquí se suelen revelar aspectos más íntimos. Por ello, servicios como *Tinder* o *Badoo* son útiles para investigar antecedentes, carácter personal, intereses, preferencias o comportamiento del target.

Finalmente, es posible navegar por comunidades en línea que son muy similares a las redes sociales. Las publicaciones y temas de estos foros generan interacciones interesantes para ser analizadas por OSINT [64]. *Reddit* o *4chan* son grandes comunidades que albergan innumerables hilos de discusión y opinión donde se puede identificar información realmente personal y privada sobre el objetivo. Sin embargo, en estos sitios web los usuarios suelen ser anónimos. Además, no es raro encontrar contenidos ilícitos de bullying, pornografía o amenazas.

Por otro lado, también hay algunas redes sociales que se utilizan normalmente en regiones específicas.
Los siguientes servicios son especialmente importantes en algunos países.

Qzone, Weibo y *Renren* son algunas de las redes sociales más utilizadas en China. La primera es una plataforma muy personalizable donde los usuarios publican blogs, diarios, fotos o música que revelan detalles sobre la persona. El segundo tiene características similares a las de *Twitter,* pero también incluye encuestas, intercambio de archivos e historias (intercambio temporal de fotos y videos). El último está muy extendido entre los estudiantes universitarios. Aquellas investigaciones OSINT cuyo objetivo sea una persona china pueden obtener una valiosa ganancia de estos sitios.

También existen redes sociales para interconectar a compatriotas rusos y ciudadanos de Europa del Este. En este sentido, *VKontakte,* también conocido como *VK,* es muy popular. Las funcionalidades, e incluso la apariencia, son bastante similares a *Facebook.* Los usuarios pueden mantenerse en contacto con amigos, participar en comunidades en línea, publicar mensajes, fotos y videos en páginas privadas o públicas e incluso compartir archivos. Otro sitio ruso a destacar es *Odnoklassniki,* utilizado principalmente por adultos.

De hecho, el objetivo principal de sus usuarios es tener un perfil en línea, mantenerse en contacto con amistades de la vida real y buscar excompañeros o antiguos amigos. En este sentido, OSINT se puede realizar para descubrir conexiones de persona a persona desde el pasado hasta el presente.

En Japón, *Mixi* es un sitio de redes sociales muy común en la sociedad. Aparte de las funcionalidades típicas, podríamos destacar la posibilidad de realizar reseñas de productos, crear blogs personales dentro de la plataforma, participar en comunidades o gestionar preferencias musicales y hábitos de escucha.

Para los países de habla hispana, especialmente Latinoamérica, *Taringa!* es una conocida plataforma social para compartir fotos, videos y noticias con amigos. Además, los usuarios pueden crear comunidades, jugar juegos en línea o compartir música.

Finalmente, debido a la censura existente con servicios externos, en Irán las redes sociales locales más populares son *Face nama* y *Cloob.* El primero se usa principalmente para compartir publicaciones, fotos y videos, mientras que el segundo incluye discusiones comunitarias, compartir fotos, publicaciones o salas de chat. Algo similar ocurre con la censura en Letonia, donde *Draugiem* es muy utilizado para compartir contenidos y comunicarse en línea.

C. TÉCNICA DE LA DIRECCIÓN DE CORREO ELECTRÓNICO

La búsqueda por el nombre real de una persona puede ser frustrante debido a la posibilidad de nombres duplicados, por lo que a veces vale la pena comenzar con una dirección de correo electrónico que sea única y logre resultados mucho mejores a un ritmo más rápido. Hay algunos servicios OSINT interesantes, como se muestra en la TABLA 4, que funcionan con una dirección de correo electrónico como entrada.

En primer lugar, *Hunter* se puede utilizar para determinar si una dirección de correo electrónico es válida o no. Luego, *Have I Been Pwned* informa si una dirección de correo electrónico determinada está contenida en infracciones públicas (de modo que se haya visto comprometida en algún momento).
En particular, vale la pena mencionar que el investigador puede navegar por la lista de sitios donde se comprometió la dirección de correo electrónico. Estos servicios son fuentes potenciales para encontrar información pública sobre el propietario. Otra página que vale la pena es *Pipl,* que funciona muy bien para encontrar información sobre el propietario de una dirección de correo electrónico como el nombre real, nombres de usuario, dirección, número de teléfono, educación, carrera profesional, etc.

D. TÉCNICA DEL NOMBRE DE USUARIO

Los apodos utilizados para los servicios en línea también son una buena manera de recopilar información sobre una persona, como se muestra en la TABLA 5. Visitar estos servicios permitirá que un investigador verifique automáticamente un nombre de usuario en varios sitios web al mismo tiempo para identificar más fuentes de información.

Los servicios *KnowEm, Name Chk, Name Checkr* o *User Search* verifican la presencia de un nombre de usuario dado en las redes sociales y dominios más populares.

NameVine, a su vez, proporciona una característica interesante que ayuda cuando se trata de adivinar un nombre de usuario exacto. Concretamente, sugiere perfiles para las diez principales redes sociales que coinciden parcialmente con el nombre de usuario dado. Esta solución en tiempo real ofrece una verificación rápida de las variantes del nombre de usuario (por ejemplo, cambiando el número final del apodo) en lugar de lanzar consultas que requieren mucho tiempo repetidamente con otros servicios.

El sitio web *Lullar* utiliza un enfoque diferente. Genera automáticamente URLs para visitar el perfil de usuario en diferentes redes sociales sin comprobar si existen. Si un enlace funciona, entonces el perfil existe para esa red social, mientras que si está roto, obviamente significa lo contrario. Además de agilizar la consulta manual, la aplicación más útil sería explorar posibles nombres de usuario cuando el que tenemos es

TABLA 4. Utilidad de los servicios OSINT pertenecientes a la técnica de dirección de correo electrónico.

Email address OSINT service	URL	Main output
Hunter	hunter.io	Validity and availability
Have I Been Pwned	haveibeenpwned.com	Appearance in public data breaches
Pipl	pipl.com	Personal information about the owner

TABLA 5. Utilidad de los servicios OSINT pertenecientes a la técnica username.

Username OSINT service	URL	Main output
KnowEm	knowem.com	Presence in social networks, domains and online communities
Name Chk	namechk.com	
Name Checkr	namecheckr.com	
User Search	usersearch.org	
NameVine	namevine.com	Suggestions of alternative similar usernames
Lullar	com.lullar.com	Availability in social networks

TABLA 6. Utilidad de los servicios OSINT pertenecientes a la técnica de nombre real.

Real name OSINT service	URL	Main output
Pipl	pipl.com	Personal information
That's Them	thatsthem.com	Personal details, education, professional career, skills, locations, and relatives.
Spokeo	spokeo.com	
Fast People Search	fastpeoplesearch.com	
Nuwber	nuwber.com	
Cubib	cubib.com	
Peek You	peekyou.com	
Yasni	yasni.com	Social networks profiles
Family Search	familysearch.org	Kinship information, relatives
GENi	geni.com	
Family Tree Now	familytreenow.com	
True People Search	truepeoplesearch.com	

cuestionable o parcial. Cuando la URL inicial falla, las redes sociales a menudo enumeran usuarios similares o alternativos que se pueden usar para identificar el nombre de usuario existente completo.

E. TÉCNICA DEL NOMBRE REAL

La búsqueda de un nombre real de destino también podría arrojar buenos resultados, como se muestra en la TABLA 6. Además de las redes sociales, los servicios particulares son capaces de revelar direcciones de casas, números de teléfono, cuentas de correo electrónico, nombres de usuario, entre otros.

Podríamos destacar *a Pipl* como la web que más información devuelve dado nombre y apellido. Debido a los posibles resultados múltiples para el mismo nombre real, es posible refinar la búsqueda al incluir aspectos adicionales de la persona, como correo electrónico, teléfono, país, estado, ciudad, nombre de usuario o edad.

That's Them también ofrece una salida notable que contiene el número de teléfono, la dirección de correo electrónico, la residencia, la dirección IP asociada, la situación económica, la educación, la ocupación o el idioma. Otro servicio muy conocido es *Spokeo,* cuya versión gratuita se reduce a mostrar nombre completo, sexo, edad, ciudades y estados de residencia anteriores y familiares. La información más detallada sobre el objetivo requiere pagar una suscripción premium, que está fuera de nuestro alcance. Servicios similares serían *Fast People Search, Nuwber, Cubib* o *Peek You.*

Los servicios antes mencionados funcionan correctamente para Estados Unidos, pero si queremos aplicar OSINT a un target que vive en otro país, el uso de *Yasni* es más adecuado. Sin embargo, los resultados obtenidos son enlaces relacionados con redes sociales, direcciones y contactos personales, educación y miscelánea.

Los servicios de genealogía como *Family Search, Family Tree Now, GENi* o *True People Search* cubren otro punto de vista en las búsquedas al proporcionar información de parentesco. Descubrir los vínculos familiares de nuestro target amplía la cantidad de información que podemos desvelar, en este caso de forma indirecta.

F. TÉCNICA DE LOCALIZACIÓN

Investigar los lugares que frecuenta nuestro target puede darnos indicios de sus hábitos y contexto. También es interesante conocer la ubicación geográfica de una empresa o el lugar donde ocurrió un evento. En este sentido, las imágenes, direcciones y coordenadas GPS son datos que vale la pena obtener. La TABLA 7 muestra algunos servicios especialmente diseñados para estos propósitos

Google Maps, Wikimapia o *Bing Maps* son sitios bien conocidos para encontrar ubicaciones a partir de coordenadas GPS. Por otro lado, también es posible obtener dicha información a la inversa de un nombre de ubicación en *Coordenadas GPS.*

TABLA 7. Utilidad de los servicios OSINT pertenecientes a la técnica de localización.

Location OSINT service	URL	Main output
Google Maps	google.com/maps	Locations from GPS coordinates
Wikimapia	wikimapia.org	
Bing Maps	bing.com/maps	
GPS Coordinates	gps-coordinates.net	GPS coordinates from location
Historic Aerials	historicaerials.com	Historic images of the past
Terra Servers	terraserver.com	
Land Viewer	eos.com	

TABLA 8. Utilidad de los servicios OSINT pertenecientes a la técnica de dirección IP.

IP address OSINT service	URL	Main output
IP Location	iplocation.net	Location, domain and ISP
ViewDNS	viewdns.info	Technical network-based information
That's Them	thatsthem.com/reverse-ip-lookup	Individual or company information
I Know What You Download	iknowwhatyoudownload.com	Torrent files

Tenga en cuenta que las imágenes que ofrecen los servicios comentados se actualizan continuamente. Sin embargo, nos podría interesar recuperar imágenes antiguas de situaciones pasadas. *Historic Aerials, Terra Servers* o *Land Viewer* incorporan funcionalidades de imágenes históricas para descubrir con precisión vistas pasadas y obsoletas de ubicaciones.

G. TÉCNICA DE DIRECCIÓN IP

Las direcciones IP se obtienen a partir de investigaciones de ciberataques, direcciones de correo electrónico o conexiones a través de Internet. También son cruciales para el análisis forense digital con el fin de recopilar la mayor cantidad de información posible de un incidente.
La TABLA 8 resume algunos servicios que facilitan estas tareas.

El servicio *de Localización IP* obtiene, a partir de una determinada dirección IP, aspectos de alto nivel como localización (latitud y longitud), país, región, ciudad, nombre de dominio o ISP *(Proveedor de Servicios de Internet).* Si estamos interesados en hechos específicos, el sitio web *ViewDNS* proporciona más información técnica además de la ubicación de la IP. En particular, incluye servicios para mostrar información de registro sobre el nombre de dominio asociado, mostrar dominios adicionales alojados en la dirección IP, descubrir puertos comunes que pueden estar abiertos y servicios ejecutándose en ellos, o ver la ruta de red desde *ViewDNS* a la IP de destino . abordar y analizar redes, enrutadores y servidores asociados.

No obstante, los recursos anteriores proporcionan datos que no tienen carácter sensible ni personal. Por el contrario, *That's Them* sí ofrece información interesante sobre personas, domicilios particulares, empresas o direcciones de correo electrónico relacionadas con la dirección IP dada.

Otro poderoso servicio que brinda información personal es *Sé lo que descargas.* Este servicio monitorea los torrents en línea y revela los archivos asociados con las direcciones IP recopiladas. Los archivos descargados por nuestro objetivo podrían revelar información realmente sensible sobre su comportamiento o intereses.

H. TÉCNICA DEL NOMBRE DE DOMINIO

Un punto típico de interés en las investigaciones OSINT son las páginas web. Pueden revelar información interesante sobre nuestro target, especialmente si se trata de una persona o una empresa. Vale la pena señalar que la mayoría de las técnicas que se explican para las direcciones IP también son adecuadas en este contexto. Además de ellos, podemos destacar algunos otros servicios como se presenta en la TABLA 9.

DNS Trails extrae registros DNS, pero también identifica la cantidad de dominios adicionales relacionados con los resultados encontrados. En este sentido, es una forma muy útil de encontrar relaciones y conexiones. *Whoisoly* también muestra una vista de referencia cruzada del nombre del propietario, dirección, número de teléfono o dirección de correo electrónico.

Otro servicio potente es *Wayback Machine,* que periódicamente realiza copias de seguridad de muchos sitios web de todo Internet. Esto le permite a un investigador analizar la evolución y los cambios de un sitio web, pudiendo verlo para capturas de pantalla particulares fechadas en el tiempo.

Además, es posible visualizar conexiones de dominio a través *de Visual Site Mapper* o *Threat Crowd.* La verificación de DNS y servidores de correo también es útil visitando *Whois,* que también ofrece una función de ping para verificar la conectividad y una función de rastreo de ruta para estudiar la ruta de datos al dominio dado. También existen servicios como *Alexa* y *SimilarWeb* que calculan estadísticas de tráfico y otros como *FindSubdomains* que buscan subdominios.

VI. HERRAMIENTAS

OSINT El uso manual de algunas técnicas sería suficiente para búsquedas básicas. Desafortunadamente, el uso de algunos servicios puede no ser efectivo para cuestionar las investigaciones. En este sentido, el potencial de OSINT radica en utilizar tantos servicios como sea posible de forma concatenada. Seguir los flujos de trabajo repetidamente ampliará la información disponible para unir todas las piezas del rompecabezas. Sin embargo, no es práctico para

TABLA 9. Utilidad de los servicios OSINT pertenecientes a la técnica de nombres de dominio.

Domain name OSINT service	URL	Main output
DNS Trails	securitytrails.com/dns-trails	DNS records and related domains
Whoisoly	whoisology.com	Personal or company information
Wayback Machine	web.archive.org/web	Backups of websites
Visual Site Mapper	visualsitemapper.com	Map of subdomains
Threat Crowd	threatcrowd.org	
Whois	who.is	Registration info and DNS records
Alexa	alexa.com	Traffic statics
SimilarWeb	similarweb.com	
FindSubdomains	findsubdomains.com	Subdomains

TABLA 10. Principales características de las herramientas OSINT seleccionadas.

OSINT tool	Input				Output	Extensibility	Interface	Platform	Other feature
	Identity data	Network data	File data	Selectable data source					
FOCA	✗	Domain	File name, Folder	Google, Bing, DuckDuckGo	Identity info, Network info, File info	✗	Stand-alone program	Windows	Server discovery module
Maltego	Personal information, company, community	Domain	File URL	✗	Identity info, Network info, File info	Custom transforms	Stand-alone program	Linux, Windows, MAC	Location, Auto input/ output refeed, Results in oriented graph
Metagoofil	✗	Domain	File type	✗	Network info, File info	✗	Command line	Linux, Windows	Option to narrow results
Recon-NG	Personal information	Domain	✗	Several	Identity info, Network info, File info	✗	Command line	Linux	Location, Modules for discovery and exploitation
Shodan	Country, City, Keyword	Operating system, IP Address, Port, Host name	✗	✗	Network info	✗	Web interface	Online	Location, Webcam captures
Spiderfoot	Email, Real name, Phone Number	Domain, IP Address, Subnet, Host name	✗	Several	Network info	Custom modules	Web interface	Linux, Windows, MAC	Different types of scan, Results in oriented graph
The Harvester	Company	Domain, DNS server	✗	Several	Identity info, Network info	✗	Command line	Linux, Windows, MAC	Results in reports, Option to narrow files and results
IntelTechniques	Personal information, company, community	Domain, IP Address	File name, File type, File URL	Several	Identity info, Network info	✗	Web interface	Online	Location, Public records, OSINT virtual machine

el usuario final para combinar manualmente varias técnicas OSINT y sus servicios asociados. Una tarea tan tediosa implicaría largos procesos de investigación.

Para ello, los investigadores y desarrolladores han implementado herramientas más precisas para aplicar técnicas OSINT automáticamente y recopilar información de mejor calidad de muchas fuentes diferentes, implementando varios flujos de trabajo internamente y, como consecuencia, obteniendo información más gratificante y mejores inferencias.

La TABLA 10 presenta las principales características de las herramientas OSINT más populares y relevantes en la actualidad. Indicamos el tipo de entradas y salidas que permiten, la capacidad de incluir funcionalidades personalizadas, el tipo de interfaz de usuario, la plataforma de funcionamiento y otras características misceláneas interesantes.

Sin embargo, hay muchas aplicaciones OSINT en el literatura a la que se puede acceder en el *marco OSINT.*[4]

A.FOCA

La principal aportación de *FOCA5 (Fingerprinting Organisations with Collected Archives),* diseñado por *ElevenPaths,* es la extracción y análisis de los metadatos presentes en los documentos electrónicos. Esta aplicación se puede utilizar tanto para archivos locales presentes en nuestro ordenador como para documentos externos que se descargan de una página web específica utilizando tres motores de búsqueda diferentes *(Google, Bing* y *DuckDuckGo).*
FOCA considera una amplia variedad de formatos como Microsoft Office, PDF, Open Office, Adobe InDesign, archivos SVG, etc.

Esta aplicación extrae la información oculta de los archivos y los procesa para mostrar al usuario aspectos relevantes. Algunos de los detalles que se descubren con este procedimiento son el nombre de las computadoras relacionadas con los documentos, la ubicación donde se crearon los documentos, los sistemas operativos utilizados,

4osintframework.com

5https://www.elevenpaths.com/es/labstools/foca-2

nombres reales y direcciones de correo electrónico de usuarios relacionados, datos sobre los servidores, fecha de creación de los documentos, rango de direcciones IP de redes internas, etc. Como resultado, se puede dibujar un mapa de red basado en los metadatos extraídos para reconocer el objetivo .

FOCA incluye además un módulo de descubrimiento de servidores para complementar el análisis de metadatos de los documentos. Algunas técnicas utilizadas en esta herramienta son: (i) *Búsqueda web* para buscar hosts y nombres de dominio a través de URL asociadas al dominio dado; (ii) *Búsqueda de DNS* para descubrir nuevos hosts y nombres de dominio a través de los servidores NS, MX y SPF; (iii) *Resolución IP* para obtener las direcciones IP de los hosts encontrados a través del DNS; (iv) *Escaneo PTR* para encontrar más servidores en un segmento de red descubierto; (v) *Bing IP* para extraer nuevos nombres de dominio asociados a las direcciones IP encontradas.

Esta herramienta se suele utilizar en el sector de la seguridad ya que permite realizar pentesting a una empresa. De hecho, es capaz de dar muy buenos resultados porque las empresas no suelen limpiar los metadatos de los archivos que suben a la red.

B. MALTEGO

Maltego6 es una conocida aplicación que encuentra automáticamente información pública sobre un objetivo determinado dentro de diferentes fuentes (registros DNS, registros Whois, motores de búsqueda, redes sociales, varias API en línea, metadatos de archivos, etc.). Las relaciones entre los elementos de interés encontrados se representan en forma de gráfico dirigido para su análisis. Esta herramienta define cuatro conceptos principales:

- **Entidad:** es un nodo del gráfico que representa la información descubierta. Algunas entidades predeterminadas son el nombre real, la dirección de correo electrónico, el nombre de usuario, el perfil de la red social, la empresa, la organización, el sitio web, el documento, la afiliación, el dominio, el nombre DNS, la dirección IP, etc. Además, también podríamos definir entidades personalizadas para nuestra investigación específica. • **Transformar:** es una pieza de código que se aplica a una entidad para descubrir una nueva entidad vinculada. Por ejemplo, la transformación *"To IP Address",* que resuelve un nombre DNS en una dirección IP, podría aplicarse a una entidad de nombre de dominio "um.es" para crear una nueva entidad de dirección IP "155.54.212.103' '. Recursivamente, seguiríamos aplicando más transformaciones, propagando el proceso de búsqueda. Además de las transformaciones predeterminadas, también es posible implementar e incluir transformaciones personalizadas para propósitos más específicos. • **Máquina:** es un conjunto de transformadas que se definen en conjunto para ser ejecutadas con el fin de automatizar y concatenar largos procesos de búsqueda.

- **Hub Item:** es un grupo de transformaciones y tipos de entidades que se utilizan para permitir que los usuarios de la comunidad los reutilicen. De forma predeterminada, Maltego implementa el elemento central llamado "Paterva CTAS" que contiene las entidades, transformaciones y máquinas mantenidas por desarrolladores oficiales.

6https://www.paterva.com/web7/buy/clientes-maltego.php

Además, es posible crear e instalar elementos de hub de terceros.

C. METAGOFILO

Metagoofil7 funciona de manera similar a FOCA. Es una herramienta de recopilación que descarga archivos públicos que se encuentran en un dominio o URL de destino y extrae sus metadatos para generar conocimiento. Genera un informe útil para pentesters con nombres de usuario, nombres reales, versiones de software y servidores o nombres de máquinas. También puede encontrar otros documentos que podrían contener nombres de recursos.

Aunque es una funcionalidad de línea de comandos, se permiten algunas opciones interesantes a favor de las investigaciones OSINT. Además de especificar el dominio de destino o la carpeta local a analizar, *Metagoofil* permite filtrar tipos de archivo (pdf, doc, xls, ppt, odp, ods, docx, xlsx, pptx), acotando los resultados a buscar y la cantidad de documentos a descargar. , determinando el directorio de trabajo donde se guardan los archivos descargados, o seleccionando el archivo para escribir la salida.

D. RECON-NG

Recon-NG8 es un marco de reconocimiento web similar a Metasploit.9 Presenta una interfaz de línea de comandos que permite seleccionar un módulo para usar, que es esencialmente un recurso OSINT. Luego, establecemos algunos parámetros si es necesario y lanzamos el proceso. Los resultados de las búsquedas se guardan continuamente en un espacio de trabajo que, a su vez, alimenta las próximas rondas del proceso.

Esta herramienta incluye varios módulos independientes que implementan diferentes funcionalidades. Por ejemplo, los módulos *Bing Domain Web* y *Google Site Web* buscan en los motores de búsqueda *Bing* y *Google* respectivamente hosts conectados a los dominios del espacio de trabajo; *PGP Search* escanea los dominios almacenados para encontrar direcciones de correo electrónico asociadas con claves PGP públicas; *Full Contact* reúne a los usuarios y los perfiles de las redes sociales correspondientes en su base de datos considerando los contactos almacenados; o *Profiler* busca servicios en línea adicionales que posean cuentas con los mismos nombres de usuario que los del espacio de trabajo.

Recon-NG está continuamente aglutinando en una base de datos local toda la información obtenida. De esta forma, el usuario dirige la investigación seleccionando el módulo indicado y la herramienta automatiza la generación de conocimiento a partir de ahí. El sistema escala notablemente para investigaciones complejas.

E. SHODAN

Shodan10 es un motor de búsqueda que proporciona información pública de los nodos conectados a Internet, incluidos los dispositivos IoT. Esto incluye servidores, enrutadores, dispositivos de almacenamiento en línea, cámaras de vigilancia, cámaras web o sistemas VoIP, entre otros. La recolección de datos se realiza a través de protocolos como HTTP o

7https://github.com/laramies/metagoofil
8https://bitbucket.org/LaNMaSteR53/recon-ng/wiki/browse
9https://www.metasploit.com/ 10https://www.shodan.io

SSH, que permite al usuario buscar por dirección IP, organización, nombre de país o ciudad.

Esta herramienta se utiliza principalmente para la seguridad de la red (para encontrar dispositivos expuestos al exterior o detectar vulnerabilidades de servicios disponibles públicamente), Internet de las cosas (para monitorear el uso creciente de dispositivos inteligentes y su ubicación en la geografía mundial) y rastrear ransomware (para medir la infección provocada por este tipo de ataques). Permite descargar los resultados en formato JSON, CSV o XML, así como generar informes amigables.

Además de la funcionalidad mencionada, hay dos servicios premium, a saber: *Shodan Maps* (maps.shodan.io), que permite investigaciones basadas en ubicaciones, y *Shodan Images* (images.shodan.io) que muestra imágenes recopiladas de dispositivos públicos.

F. PATA DE ARAÑA

Spiderfoot11 es otra herramienta de reconocimiento que revisa automáticamente muchas fuentes de datos públicos para recopilar información. Nuestra entrada podría ser una dirección IP, subred, nombre de dominio, dirección de correo electrónico, nombre de host, nombre real o número de teléfono. Los resultados se representan en un gráfico de nodos con todas las entidades y relaciones encontradas.

Dependiendo del tipo de entrada introducida, esta herramienta selecciona de forma autónoma los módulos (equivalentes a las transformadas de Maltego) para activar para un reconocimiento más efectivo.
Además, también considera el nivel de búsqueda seleccionado por el usuario. *Spiderfoot* ofrece cuatro tipos de escaneos: (i) *Pasivo* recopila la mayor cantidad de información posible sin tocar el sitio objetivo, evitando que el objetivo lo descubra; (ii) *Investigar* realiza un escaneo básico para descubrir la malicia del objetivo; (iii) *Footprint* identifica la topología de red del objetivo y recopila información de la web y los motores de búsqueda, suficiente para investigaciones estándar; y (iv) *Todos,* lo cual es recomendable para investigaciones detalladas, a pesar de llevar mucho tiempo en completarse, ya que consulta absolutamente todos los recursos posibles relacionados con el objetivo.

Esta herramienta podría usarse para lanzar pruebas de penetración para revelar fugas de datos y vulnerabilidades, desafíos de equipo rojo o para respaldar la inteligencia de amenazas. Además, vale la pena señalar que es posible programar módulos *Spiderfoot* personalizados.

G. EL COSECHADOR

El Harvester12 permite la recopilación de información pública relacionada con un dominio o nombre de empresa a través de motores de búsqueda.
En particular, es capaz de listar correos electrónicos y nombres de host de la empresa, así como subdominios, direcciones IP y URL relacionadas con el dominio. También permite representaciones HTML o XML fáciles de usar de los resultados. Este recurso se utiliza en las primeras etapas de una prueba de penetración.

Esta herramienta se gestiona desde la consola e implementa dos opciones a la hora de escanear nuestra web de destino. Por un lado, *The Harvester* representa el guión original que en realidad proporciona la lista de direcciones de correo electrónico relacionadas, mientras que, por otro lado, *EmailHarvester* mejora el procedimiento profundizando para obtener mejores resultados.

H. INTELTECHNIQUES

IntelTechniques13 es una herramienta, creada por Michael Bazzel, que ofrece cientos de utilidades de búsqueda en línea agrupadas por técnica.

Al utilizarla, el investigador selecciona los servicios a utilizar y esta herramienta crea automáticamente los enlaces de consulta asociados. Posteriormente, el usuario puede ingresarlos en el navegador para lanzar las consultas. Sin embargo, la visualización y recopilación de la información sigue siendo manual.

A pesar de que no implementa una integración automática de servicios, hemos considerado a *InterTechniques* como una herramienta OSINT que facilita el lanzamiento de búsquedas a una amplia gama de servicios desde una plataforma centralizada.

Lamentablemente, esta herramienta dejó de ser gratuita y bloqueó su acceso abierto a partir de julio de 2019 debido a los constantes ataques cibernéticos.

I. COMPARACIÓN DE HERRAMIENTAS OSINT

Dependiendo de las necesidades del usuario (ver TABLA 10), algunas herramientas serán más adecuadas que otras para una tarea determinada.

Así, si pretendemos extraer **información oculta de los archivos,** *FOCA* y *Metagoofil* son herramientas específicas diseñadas para tal fin. En particular, el primer producto parece ser más completo, maduro y potente que el segundo. *FOCA* presenta funcionalidades adicionales, además del análisis de metadatos de archivos, para complementar la información oculta.

Como resultado, es capaz de inferir más conocimiento sobre el objetivo.

Sin embargo, si buscamos **información de la red,** *Shodan, Spiderfoot* y *The Harvester* son opciones recomendadas para esta determinada tarea. Por un lado, sugeriríamos que *Spiderfoot* analice la topología del objetivo y recupere información interna (pero pública) sobre la organización objetivo. Por otro lado, completaríamos los resultados con *Shodan* para incluir información específica sobre dispositivos IoT, cámaras de vigilancia, webcams, sistemas VoIP o servicios inteligentes en general.

Por último, pero no menos importante, si el objetivo de la búsqueda es recopilar **la mayor cantidad de información posible** para una determinada entrada, los recursos *Recon-NG* y *Maltego* son los más completos y devolverán diversos datos y relaciones.
El primero contiene gran cantidad de módulos e interactúa con una base de datos local que escala durante la investigación, siendo un marco ideal para realizar pentestings, prevención de ataques de phishing e ingeniería social, o incluso la elaboración de perfiles de una persona. Por el contrario, si queremos evitar la línea de comandos y optar por una interfaz más amigable, *Maltego* es una buena alternativa para las actividades OSINT. Implementa procesos de inferencia automatizados con transformadas que elevan el alcance de la búsqueda original. Además, es extensible con procedimientos de descubrimiento personalizados.

11https://www.spiderfoot.net
12https://github.com/laramies/theharvester
13https://inteltechniques.com

A pesar de que la comparación descrita anteriormente se ha realizado de acuerdo con la salida deseada, en la práctica el usuario estará restringido por la entrada disponible y el tipo de datos aceptado por las herramientas OSINT elegidas. Finalmente, tenga en cuenta que estas herramientas son complementarias y no excluyentes entre sí, lo que significa que una investigación OSINT profunda y exhaustiva podría beneficiarse de varias de ellas al mismo tiempo. Aunque algunos de ellos pueden producir resultados similares para una búsqueda determinada, siempre puede haber detalles encontrados por una herramienta en particular que no son obtenidos por otros.

VIII. INTEGRACIÓN DE OSINT EN CIBERATAQUE INVESTIGACIONES

La implementación de mecanismos de detección y respuesta a los ciberincidentes es hoy una obligación. Las empresas y organizaciones, cada vez más expuestas en Internet, invierten en ciberseguridad para proteger sus activos frente a los delincuentes. Por lo tanto, es de suma importancia gestionar las amenazas e incidentes contra los sistemas de información de manera efectiva.

La ciberdefensa no es sólo el despliegue de soluciones técnicas como cortafuegos, IDS (Sistemas de *Detección de Intrusos),* IPS *(Sistemas de Prevención de Intrusos),* SIEM *(Security Information and Event Management)* o antivirus para evitar amenazas conocidas, sino también la implantación de ciberinteligencia para extraer y analizar rastros, patrones y conclusiones de los incidentes. De hecho, el ciclo continuo de extraer y compartir evidencias, relaciones y consecuencias de incidentes se conoce como inteligencia de amenazas [65]. Complementa los mecanismos de defensa tradicionales con información actualizada y mejora notablemente la protección de las infraestructuras, la gestión de los peligros y la eficacia de las respuestas [41].

Además, la información que normalmente se utiliza para el análisis forense y las investigaciones es meramente técnica. Sin embargo, los rastros que deja un ciberataque contienen información valiosa que no solo debe contrastarse con repositorios de incidentes [66], sino también con redes sociales, foros, medios de comunicación, documentos técnicos y gubernamentales y otras fuentes públicas digitales. Estas fuentes abiertas aportan información semántica en el análisis, lo que resulta interesante para computar y razonar inferencias más complejas y de mayor alcance. Tenga en cuenta que los ciberatacantes utilizan Internet para sus acciones ilegales (piratería, phishing, ataques de denegación de servicio, botnets, robo de identidad, intrusiones, etc.), pero también por motivos personales. En este sentido, OSINT se puede utilizar para conectar todos esos puntos.

Varios trabajos de aplicación de OSINT a la ciberseguridad se centran en proponer mejoras defensivas ante amenazas.
Por el contrario, muy pocas veces buscan la identificación de los ciberatacantes. OSINT es una fuente de conocimiento que podría apoyar la investigación de un ciberataque al ir desde los detalles más pequeños de la acción maliciosa hasta la raíz del problema. Este último desafío no es nuevo, ya que tradicionalmente se conoce como problema de atribución [67]. Concretamente, OSINT nos permitiría comprender la motivación de la ciberataque, adivinar el procedimiento y, en última instancia, perfilar al perpetrador.

La aplicación sugerida de OSINT se ilustra en la FIGURA 3. Tenga en cuenta que se han propuesto varias metodologías y modelos para definir la madurez de detección de una organización, que es crucial para extraer evidencias de un ciberataque sufrido. No obstante, faltan estándares para representar taxonomías y ontologías en este campo [68], por lo que proponemos una versión modificada del modelo DML de Ryan Stillions [69] para ejemplificar esta sección. Sin embargo, se podría usar otro esquema de detección de amenazas cibernéticas para mostrar la aplicación de OSINT de manera similar.

El modelo DML representa de forma jerárquica diferentes niveles de abstracción en la detección de ciberataques. Una empresa que no invierte en ciberseguridad solo podrá llegar a los escalones más bajos de la pila. Por el contrario, una organización técnicamente experta en ciberdefensa puede interpretar hechos más complejos, es decir, ascender a niveles de mayor abstracción.

Si bien los niveles inferiores se pueden cubrir fácilmente, el desafío radica en llegar a las capas superiores. Para ello sugerimos aplicar OSINT como fuente de inteligencia que se alimenta de la evidencia más básica para llegar a hechos más robustos:

1) En primer lugar, suponemos que es posible cubrir los niveles DML-1 y DML-2. El primero, *Indicadores atómicos de compromiso (IOC),* está compuesto por detalles tan simples como una cadena en un archivo modificado, el valor de una celda de memoria o un byte transmitido a través de la red, que tienen un valor muy bajo por sí mismos, pero juntos forman el siguiente nivel. La capa *de artefactos de host y red* se basa en los indicadores observados durante o después del ciberataque, como direcciones IP, nombres de dominio, registros, transacciones, valores hash o detalles de manipulación de archivos.

   Como este tipo de datos reside en los sistemas de información afectados, en nuestro marco se considera como un insumo para la recopilación de información asociada en fuentes abiertas (consulte la SECCIÓN V para obtener más detalles sobre la recopilación de OSINT). Por lo tanto, la extracción de estos rastros es el punto de partida de un proceso OSINT.
2) A continuación tenemos desde el nivel DML-3 hasta el nivel DML-6. Las *Herramientas* de tercer nivel consisten en detectar la transferencia, presencia y funcionalidad de las herramientas utilizadas por el atacante. El siguiente nivel de *Procedimientos* está cubierto si uno es capaz de enumerar los pasos realizados durante el incidente.
   Las *Técnicas* de quinto nivel extraen cómo el atacante ha realizado específicamente las diversas fases del ataque.
   Y el último nivel aquí, *Tácticas,* es un concepto más abstracto que tiene en cuenta los niveles discutidos anteriormente y deriva conocimiento al analizar un conjunto de actividades en el tiempo y el contexto.

   En este caso, la información revela detalles sobre la ejecución del ciberataque. Dichos datos enriquecen enormemente la fase de análisis del ciclo OSINT. Los patrones derivados de estos datos, así como la correlación con otros casos ya almacenados, nos permiten tener un análisis más inteligente y completo. De hecho, estos

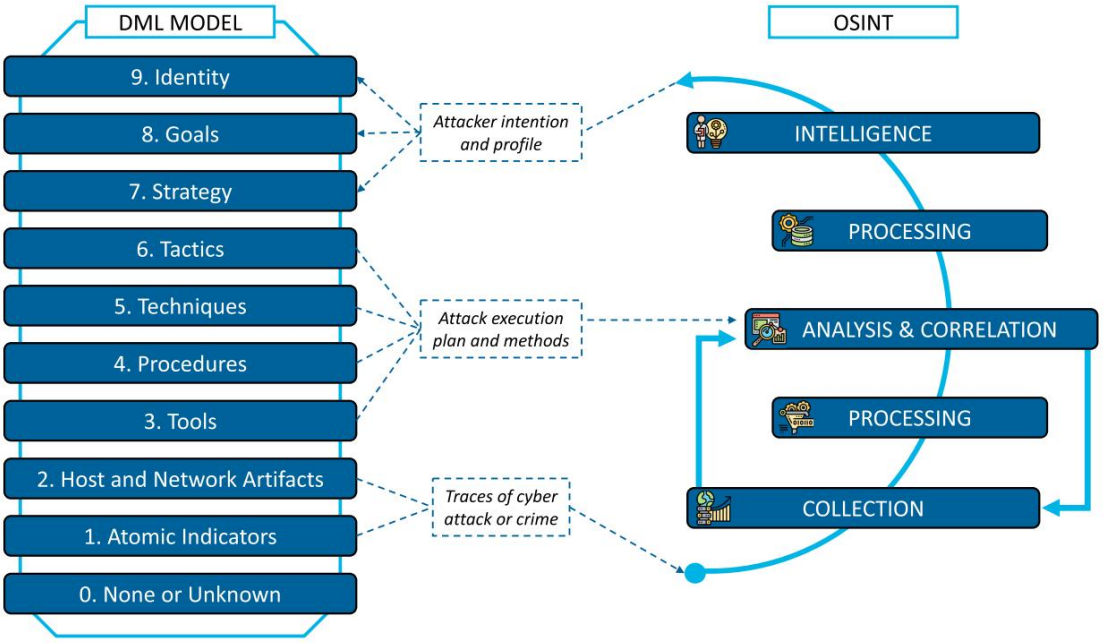

FIGURA 3. Integración de OSINT con el modelo DML para abordar el problema de atribución.

las conclusiones deben integrarse junto con los resultados obtenidos en la fase de recopilación. De esta forma se afina la exploración a través de la red, acotando la investigación hacia el objetivo final.

3) Finalmente, el proceso continuo de recopilación y análisis de OSINT genera información valiosa a la que se aplican técnicas de extracción de conocimiento.
El conocimiento extraído con OSINT del nivel DML-1 al DML-6 nos permitiría llegar a los niveles más altos, es decir, del DML-7 al DML-9. El séptimo nivel, *Estrategia,* se refiere a una descripción de alto nivel del ataque planeado por el ciberdelincuente para completar sus propósitos. El octavo nivel, *Goals,* son los objetivos específicos del atacante y expresan la motivación real de la acción. En la parte superior encontramos el nivel de *Identidad ,* que es esencialmente el nombre de una persona, una organización o incluso un país responsable de las acciones maliciosas. Como es extremadamente difícil encontrar esa información detallada, la conexión con otros ciberataques y la similitud con otros eventos pueden respaldar la atribución relativa [67]. Es decir, completar la investigación del caso actual con información adicional sobre otros incidentes aparentemente provocados por el mismo actor nos acerca a la identificación absoluta del ciberatacante.

Esta aplicación de OSINT representa una innovadora línea de actuación para luchar contra las ciberamenazas. El reto reside en implementar mecanismos efectivos de recolección y procedimientos de análisis inteligente para extraer aquellos detalles de alto nivel que no pueden ser extraídos directamente de acciones maliciosas. Dichos detalles son las piezas de información más complicadas de lograr, ya que tienen un grado de abstracción muy alto que están muy lejos de los detalles técnicos. por eso es

Es inteligente buscar en las fuentes abiertas cualquier relación o patrón que nos lleve a descubrir más sobre el contexto y los originadores de un incidente. OSINT es la pieza clave que faltaba en el equipo para perfilar a los ciberatacantes y mejorar la detección de ataques sofisticados [70] gracias a la consideración de aspectos de comportamiento de alto nivel desde DML-3 hasta DML-9.

VIII. OSINT EN PAÍSES Y ESTADOS

La OSINT no solo es beneficiosa en el sector privado, sino que también representa un recurso de interés público en los gobiernos. Al respecto, en la SUBSECCIÓN VIII-A discutimos que OSINT no es un paradigma diseñado para analistas paranoicos o informáticos, pero sí tiene un enorme beneficio en el sistema nacional de ciberdefensa [71]. Asimismo, en el INCISO VIII-B observamos que las autoridades oficiales no sólo se benefician de los resultados de OSINT para tareas internas, sino que indirectamente facilitan la aplicación de OSINT a terceros. De hecho, se convierten en un agente que genera grandes cantidades de datos accesibles para todos. En este sentido, los gobiernos son un arma de doble filo que se benefician de OSINT pero al mismo tiempo contribuyen a alimentar Internet con información realmente valiosa, ya veces incluso sensible.

A. OPERACIONES DE ASUNTOS INTERNOS DEL ESTADO

Las Agencias de Inteligencia se han asociado tradicionalmente con el trabajo de las Agencias de Aplicación de la Ley (LEA) y los Cuerpos Militares. De la misma manera, OSINT es considerado hoy en día como una clave importante de investigaciones clasificadas y operaciones secretas en asuntos de estado [5]. Hasta cierto punto, se podría argumentar con seguridad que la explotación de OSINT puede proporcionar capacidades críticas para que las LEA complementen y mejoren sus departamentos de contrainteligencia en la investigación y planificación estratégica para luchar contra el crimen [72].

Por lo que pudimos explorar en los sitios web oficiales, informes y documentación, las organizaciones gubernamentales parecen implementar mecanismos internos que consisten básicamente en recopilar información en bruto y transformarla en conocimiento útil, aprovechando los mecanismos OSINT [73]. De manera representativa podríamos mencionar a la *Oficina Federal de Investigaciones de EE. UU. (FBI,* fbi.gov), *la Agencia Central de Inteligencia de EE. UU. (CIA,* cia.gov), *el Servicio de Inteligencia de Seguridad de Canadá (CSIS,* canada. ca/en/security-intelligence -service), *Agencia de la Unión Europea para la Cooperación en materia de Aplicación de la Ley (EUROPOL,* europol.europa.eu), *Organización del Tratado del Atlántico Norte (NATO,* nato.int), *Departamento del Ejército de los Estados Unidos (DA,* army.mil), *Departamento de Defense (DoD,* defense.gov), *la Agencia de Seguridad Nacional de EE. UU. (NSA,* nsa.gov) o *la Agencia Europea de Defensa (EDA,* eda.europa.eu), entre otras.

En este escenario de incertidumbre, hemos decidido investigar en particular el caso de las LEA españolas, por afinidad, para demostrar que los organismos oficiales internamente sí aplican OSINT. Como resultado de esta minuciosa inspección, podemos afirmar enfáticamente que no es fácil encontrar evidencias claras de la aplicación de OSINT por parte de las fuerzas del Estado. La confidencialidad de este tipo de agencias dificulta conocer su modo de funcionamiento interno y el impacto de OSINT en sus investigaciones actuales. Sin embargo, como consecuencia de la búsqueda profunda, tenemos algunos hallazgos sutiles que confirman que OSINT es utilizado actualmente por las LEA españolas:

- Allá por 2007, el director del CNI (es decir, la Agencia Nacional de Inteligencia de España) dijo14 que las fuentes abiertas eran *"fundamentales para la elaboración y el trabajo de inteligencia"*
- CIFAS (es decir, la Agencia Española de Inteligencia Militar) también parece utilizar OSINT como una forma de obtener información. Hemos encontrado unas diapositivas que así lo confirman, fechadas ya en 2008, que están subidas a la web del Estado Mayor de la Defensa.15 • En 2010, cuando el director del CNI anunció16 la creación de un código ético para los agentes especiales, también insistió en que la inteligencia moderna no se basaba sólo en la presencia física, ya que hoy *"usted puede obtener más información sentado en una computadora, explorando los mensajes de los malos".*
- Más recientemente, en 2017, el Ministerio de Defensa español abrió una convocatoria pública17 para el contrato denominado *"Desarrollo de la herramienta OSINT basada en la plataforma IDOL HAVEN".* • En la actualidad, el Ejército de Tierra está diseñando un nuevo modelo denominado *Brigada 2035* que incorpora avances tecnológicos innovadores para mejorar las operaciones. En este proyecto,18 una de las funciones de combate definidas es *Inteligencia,* lo que establece claramente que OSINT es una responsabilidad clave: *"Otras instalaciones de creciente importancia serán la obtención de código abierto (incluidas las redes sociales)".*
- El Ministerio del Interior de España ha publicado en el Plan Anual de Contratación de 201919 unas inversiones en *"sistemas para la obtención de OSINT en el ciberespacio".*

Teniendo en cuenta todos estos hechos, parece que actualmente OSINT es realmente relevante en los asuntos internos de España. Análogamente, también podríamos destacar que los estados miembros de la Unión Europea también están muy desarrollados en OSINT [74].

B. POLÍTICAS DE DATOS ABIERTOS Y TRANSPARENCIA

OSINT depende de los datos públicos disponibles en Internet, entre otras fuentes, para ser efectivo. En este sentido, además de las redes sociales y otras fuentes de datos abiertos, también existen sitios autorizados y oficiales mantenidos por instituciones estatales de todo el mundo donde se publica información pública y, por lo tanto, disponible abiertamente.

El Barómetro de Datos Abiertos (ODB)20 es un sistema de clasificación global diseñado por la World Wide Web Foundation que mide la preparación, implementación e impacto de las políticas de datos abiertos de los países. En la Figura 4 se muestran las puntuaciones de la última edición completa.21

Como ya hemos hecho en el apartado anterior, estudiamos el caso concreto de España por afinidad. De hecho, respecto al citado informe ODB, España se sitúa en el puesto 11. Además, según el Portal Europeo de Datos y sus informes oficiales22 sobre la madurez de los Datos Abiertos en Europa, España es uno de los países más avanzados en transparencia y datos abiertos. Ha estado en primera o segunda posición en el ranking de Madurez de Datos Abiertos en los últimos cuatro años. Tal y como se afirma, el Gobierno de España ha impulsado más de 160 iniciativas de datos abiertos y cuenta con más de 23.800 catálogos de información pública. Por ejemplo, la Iniciativa de Datos Abiertos del Gobierno de España23 es una muestra clara de cómo España fomenta la transparencia. OSINT podría beneficiarse de eso, pero debería tratar con información agregada y estadística vinculándola e infiriendo nuevos conocimientos.

También existen bases de datos anonimizadas que, a priori, no serían útiles para OSINT porque carecen del valor para producir inteligencia. Aparentemente, estos conjuntos de datos llamados anónimos no rompen el vínculo entre los datos y su propietario.
Recientemente, se ha publicado un algoritmo [75] que permite identificar inequívocamente al 99,98% de los estadounidenses a partir de datos públicos. En particular, basta con tener 15 parámetros relacionados con los aspectos médicos, conductuales y sociodemográficos.

14https://www.elconfidencialdigital.com/articulo/vivir/CNI-califica fundamental-abiertas-contradice/20071023000000049386.html
15http://www.emad.mde.es/Galerias/EMAD/novemad/fichero/EMD CIFAS-esp.pdf 16https://www.lavanguardia.com/politica/ 20100624/53951898847/el director-del-cni-anuncia -un-codigo-etico-para-los-agentes-secretos.html 17https://contrataciondelestado.es/wps/ wcm/connect/ff96fa82-7fd6- 40bd-be5b-36ef3fd4e65b/ DOC_CN2017-498874.pdf?MOD=AJPERES
18www.ejercito.mde.es/es/estructura/briex_2035/principal.html
19http://www.defensa.gob.es/Galerias/gabinete/ficheros_docs/2019/ PACDEF_2019_Documento_Pxblico.pdf 20https ://opendatabarometer.org
21https:/ /opendatabarometer.org/4thedition 22https:// www.europeandataportal.eu/es/dashboard#2018 23https:// datos.gob.es/es

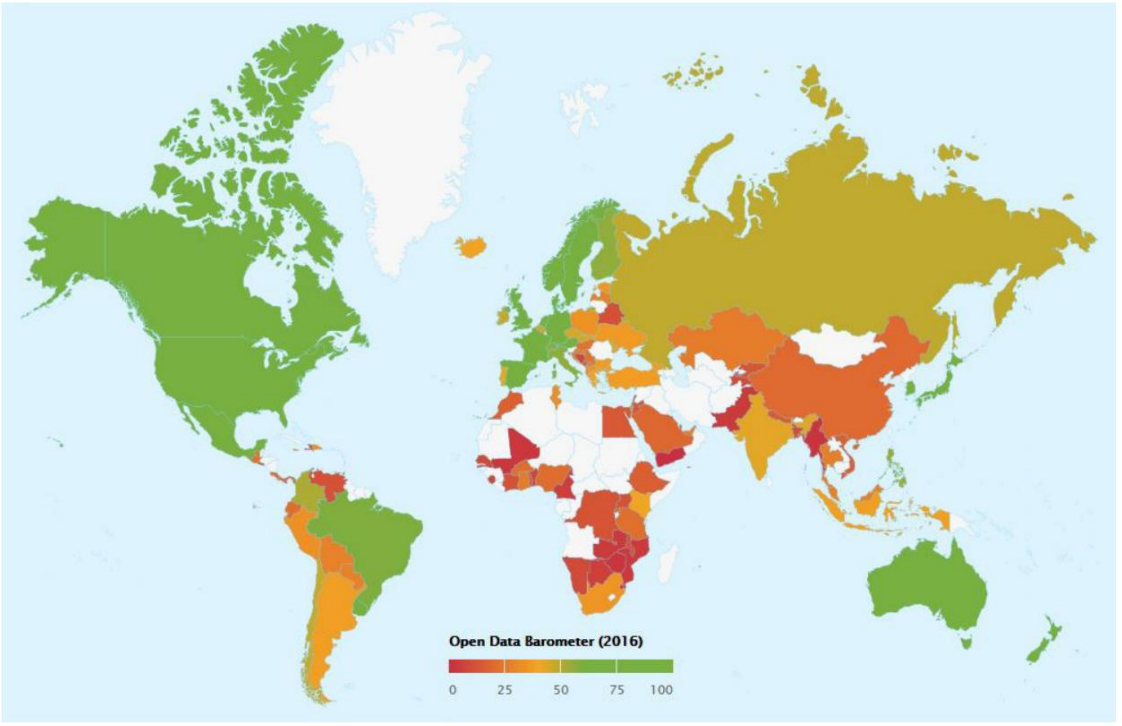

FIGURA 4. Puntuaciones de transparencia de la 4ª edición del Barómetro de Datos Abiertos.

información como estado civil, sexo o el código postal de su domicilio. Por lo tanto, OSINT podría usarse nuevamente para volver a identificar a las personas recopiladas en bases de datos anónimas.

Por el contrario, también existen plataformas gubernamentales que en realidad no son anonimizadas. Por ejemplo, el Ministerio de Hacienda español, el Ministerio del Interior español o el Ministerio de Defensa español suelen publicar documentos con información personal ("site:hacienda.gob.es filetype:pdf intext:dni", por ejemplo). De la misma manera, esto podría aplicarse también a los sitios web de las Comunidades Autónomas españolas.

Además, Europa también tiene una plataforma de datos públicos24 , donde podemos encontrar mucha información pública. Por ejemplo, en el contexto de la política exterior y la seguridad, se presenta una lista actualizada de sanciones financieras en el documento *"Lista consolidada de sanciones financieras de la Unión Europea"* . En particular, revela información personal sobre individuos, grupos y entidades.

Todos los hechos mencionados anteriormente demuestran que los gobiernos de todo el mundo están adoptando fuertes políticas de datos abiertos. Como consecuencia directa, la cantidad de datos objetivos disponibles en Internet está aumentando rápidamente. OSINT debe, además de otras fuentes abiertas de información, aprovechar esta poderosa oportunidad para recopilar, analizar, vincular e inferir conocimientos de fuentes confiables y oficiales. En este escenario, y según el ODB, son reales países como Reino Unido, Canadá, Francia, Estados Unidos, Corea, Australia, Nueva Zelanda, Japón, Países Bajos, Noruega o Brasil.

24http://data.europa.eu/euodp/en/data

Minas de oro OSINT de características muy similares a las comentadas para España.

IX. RETOS ABIERTOS Y TENDENCIAS FUTURAS

La revisión realizada sobre OSINT muestra que ya existe una cantidad sustancial de trabajo en el tema. Numerosas técnicas y herramientas se han desarrollado hasta ahora. Sin embargo, existen algunos vacíos y limitaciones en este campo para seguir explotando las oportunidades que se ofrecen. Es necesario realizar soluciones más sofisticadas aplicables a escenarios no controlados del mundo real. Hemos detectado algunos desafíos que, hasta donde sabemos, están abiertos en la actualidad y deberían ser enfrentados por la comunidad investigadora en el futuro próximo.

A. AUTOMATIZACIÓN DEL PROCESO DE RECOLECCIÓN

Cuanto mayor sea la cantidad de información recopilada, más probable es que se creen inferencias y relaciones. Sin embargo, la cantidad de datos públicos disponibles hoy en día es enorme y no se puede recopilar de forma manual [76]. Aunque las técnicas OSINT (Sección V) y las herramientas (Sección VI) ya son un gran paso adelante en esta dirección, la mayoría de ellas todavía dependen en gran medida del usuario final. En este sentido, sería atractivo incorporar técnicas más sofisticadas. Destacamos las técnicas actuales de big data como Web crawling o Web scraping [77] como paradigmas potenciales para automatizar y mejorar la exploración OSINT de grandes volúmenes de datos abiertos.

Un aspecto importante del proceso de recolección es la propagación de la búsqueda. Los resultados obtenidos con las búsquedas deberían realimentar las siguientes rondas de recolección.
En OSINT es realmente poderoso extraer pivotes que permitan

la concatenación de salidas como nuevas entradas para la propagación. Este método recursivo aumenta el alcance de la investigación y está estrechamente relacionado con el proceso de análisis que discutiremos a continuación.

B. MEJORA DEL ANÁLISIS Y EL CONOCIMIENTO PROCESOS DE EXTRACCIÓN

La interpretación de los datos abiertos recolectados es un punto clave en el procedimiento OSINT. Extraer la esencia de los resultados del scraping, establecer relaciones entre piezas de información separadas o inferir conclusiones que no están explícitamente expuestas aumenta la calidad de los resultados. De hecho, la integración recursiva con la propagación de nuevas rondas de investigación se ve reforzada por medio de mejores entradas.

Sin embargo, hasta donde sabemos, el análisis OSINT no está implementando mecanismos inteligentes en la actualidad. Las herramientas existentes se limitan a arrojar toda la información encontrada y sus relaciones explícitas. Por el contrario, el proceso de análisis debe incorporar análisis semántico, estudio de patrones, correlación con otros eventos, ocurrencias o conjuntos de datos.

Afortunadamente, las técnicas modernas de minería de datos [78], como el procesamiento del lenguaje natural, el análisis de redes sociales, el aprendizaje automático o el aprendizaje profundo, están diseñadas para resolver este tipo de desafíos. Una adecuada selección de algoritmos en este campo del conocimiento marcará la diferencia entre el análisis estático actual y el futuro procesamiento razonado [79].

Idealmente, el OSINT del futuro debería ser capaz de proporcionar al usuario final la información específica que está buscando, así como devolver respuestas convincentes en las investigaciones. La búsqueda original también tendría, no sólo inferencias directas, sino también relaciones indirectas y no explícitas.

Este desafío construye el camino entre la Segunda Generación y la Tercera Generación de OSINT. Como se presenta en [1], la Segunda Generación comenzó con el surgimiento de Internet y las Redes Sociales, y los desafíos eran *"experiencia técnica, accesibilidad virtual y adquisición constante"*.
En cambio, se supone que la evolución hacia la Tercera Generación aparecerá hoy en día y deberá incluir *"procesamiento automático directo e indirecto de datos, aprendizaje automático y razonamiento automatizado"*.

C. INTEGRACIÓN DE VARIAS FUENTES DE DATOS ABIERTOS

Las actividades de OSINT deben consultar tantas fuentes como sea posible para cubrir el espectro más amplio posible. No es una buena idea centrar nuestra investigación en una sola red social o un foro específico. En este sentido, el éxito radica en combinar las fuentes de datos para obtener los mejores resultados posibles. Esto significa que el sistema tiene que normalizar la información disponible, que normalmente no está estructurada, para poder realizar un análisis y correlación efectivos. Como resultado, es importante descartar elementos repetidos. De hecho, las diferentes técnicas y herramientas OSINT que se explican en este documento están aplicando tal sesión para recopilar el conocimiento relacionado con el objetivo.

Por otro lado, el verdadero desafío es incorporar, no solo varias fuentes de datos, sino diferentes tipos de datos.

fuentes [80]. Además de los datos extraídos de Internet, Dark Web y Deep Web, el flujo de trabajo OSINT también debe considerar la información recopilada cara a cara, con ingeniería social o con la colaboración de los ciudadanos. Cualquier pieza de información que sea interesante para nuestra investigación tiene que ser utilizada para lograr el próximo hito de la búsqueda. Además, es imprescindible la implementación de procesos de descubrimiento de la verdad para aquellos casos en que la información de diferentes fuentes de datos sea contradictoria [81].

D. FILTRADO DE DATOS IRRELEVANTES Y LA DESINFORMACIÓN

Debido a la gran cantidad de datos disponibles públicamente, un proceso OSINT debe ser capaz de distinguir la relevancia de cada información, descartando datos que no agregan calidad a la investigación [82]. Un investigador no puede concentrarse en explorar los detalles de un sitio web completo, leer una noticia de varias páginas o analizar un documento gubernamental complejo. Por el contrario, la investigación de OSINT necesita extraer palabras clave que realmente aporten valor y revelen conocimiento sobre nuestro objetivo. Es posible que la información que nos interesa no se publique explícitamente, y el desafío sería extraer la esencia de la fuente de datos que estamos examinando.

Al mismo tiempo, los términos precisos extraídos sirven como pivotes para crear nuevos caminos de exploración.

Además, es crucial detectar información errónea que podría corromper los resultados [83]. Por naturaleza, Internet es subjetivo y la mayoría del contenido no tiene garantía de ser confiable y oficial. La comunidad OSINT tiene que determinar si la creciente confianza en los datos de fuente abierta todavía se combina con la validación de las fuentes, lo que representa un requisito principal y una prioridad [84]. Esa información falsa puede desviar nuestra búsqueda, llevándonos a resultados erróneos o alejados de nuestro objetivo real. Por ello, sería interesante analizar no solo la información objetiva, sino también la información falsa con el objetivo de extraer inteligencia.

Este problema estará presente en la investigación de la vida real. Las fuentes de datos donde encontraremos información más valiosa sobre sospechosos serán en foros y redes sociales. En estos sitios, el investigador tiene que lidiar con opiniones, publicaciones subjetivas y preferencias personales cuya veracidad es cuestionable [85]. La elaboración de perfiles de personas que en realidad no representan una amenaza (falsos positivos) podría provocar actitudes discriminatorias e injustas que podrían afectar a las víctimas.

E. EXTENSIÓN POR TODO EL MUNDO

Uno de los principales inconvenientes de muchos de los recursos OSINT existentes es que solo funcionan para países específicos, lo que reduce su capacidad de creación de perfiles a un grupo restringido de personas pertenecientes a unas pocas nacionalidades. Sin embargo, OSINT debería ser una técnica universal para recorrer todos los rincones de la Tierra al instante sin discriminar zonas del ciberespacio. Por lo tanto, la interoperabilidad es una propiedad deseable a considerar en el diseño de OSINT ya que aumentaría, no solo el alcance de las búsquedas, sino también su uso por parte de los usuarios finales.

Idealmente, un buen servicio o herramienta OSINT no debe distinguir entre países y tomar cada investigación como una tarea global, sin fronteras. El flujo de trabajo OSINT debe combinar puntos de información en todo el mundo y correlacionar esas fuentes de datos distribuidas. De hecho, aunque la relación entre las zonas de búsqueda podría hacerse a mano, el verdadero desafío radica en que las aplicaciones OSINT implementen estos *saltos.*

Además, la globalización del proceso no dejaría de lado atractivas fuentes de datos abiertos de diferentes territorios que en realidad podrían llenar los vacíos que debemos abordar en nuestra investigación. En España, por ejemplo, utilizamos herramientas diseñadas en (y para) países extranjeros. Sin embargo, no existen soluciones OSINT que incluyan repositorios públicos españoles en fase de recopilación (como podrían ser las plataformas de datos abiertos gubernamentales). En este sentido, todavía no nos estamos beneficiando del todo de la mina de oro que supone ser uno de los países más transparentes de Europa.

Una implementación genérica y flexible es especialmente útil para *objetivos nómadas* en los que la movilidad forma parte de su día a día. Decir que el objetivo investigado es una persona que ha vivido etapas de su vida en varios países, o empresas que tienen sedes en varios continentes, o incluso delincuentes que cambian de ubicación para dificultar su persecución. En estos casos, una búsqueda estática en un país en particular dejaría mucha información sin recopilar y muchas pistas sin analizar.

F. CONCIENTIZACIÓN DE LA PRIVACIDAD, ÉTICA Y CONSIDERACIONES LEGALES

Desde un punto de vista ético, OSINT debe respetar la privacidad del usuario para no dañar su vida privada, así como la privacidad de su familia, amigos y compañeros de trabajo. El hecho de que la información sea de acceso público no significa que no sea sensible. Conocer las preferencias y gustos personales del target puede perpetrar en su intimidad. Revelar pensamientos políticos puede tener consecuencias fatales en ciertos lugares. Comunicar una orientación sexual puede ser potencialmente mortal en ciertos países. Conocer las creencias religiosas puede dar lugar a condenas penales en territorios específicos. Por lo tanto, la información de fuente abierta debe manejarse con cuidado, para fines legítimos, en interés de la sociedad.

Desde el punto de vista legal, OSINT debe usarse sobre la base de una ley y respetando las políticas de protección de datos. Con la llegada del RGPD de la UE, la regulación relativa a los datos personales ha cambiado [86]. En este sentido, los datos personales comprenden cualquier información que pueda relacionarse con cualquier ciudadano.
Además, diferentes piezas de información, que recopiladas juntas pueden conducir a la identificación de un individuo, también constituyen datos personales, incluso si la información está encriptada o anonimizada [14]. Una posible solución para abordar este desafío es adaptar el diseño de las herramientas OSINT para incorporar restricciones normativas, especialmente los requisitos legales de GPDR [87].
Por definición, OSINT es completamente legal debido a la naturaleza pública de las fuentes de datos que utiliza. No obstante, los investigadores no debe publicar la información personal recopilada, incluso si se publica en la web. Además, el usuario que aplica OSINT no puede caer en el error de intentar suplantar al objetivo para encontrar más información. También cabe señalar que las barreras de autenticación no se pueden romper para acceder a la información que buscamos.

En resumen, el uso de OSINT debe restringirse a actividades legales y propósitos no maliciosos. En principio, OSINT no viola (y no debería violar) la libertad y los derechos humanos, por lo tanto, sus técnicas y servicios mencionados anteriormente son legales en esta medida [88]. Es una metodología realmente poderosa, pero también es peligrosa si se usa mal. Gracias a OSINT, los periodistas pueden brindar noticias actualizadas, objetivas y de calidad. Los gerentes de recursos humanos pueden conocer mejor a los solicitantes en su trabajo. Las autoridades de los países pueden investigar grupos criminales y terroristas. Una empresa puede auditar su exposición en el extranjero a las ciberamenazas. Sin embargo, dicha apertura a la utilización de técnicas OSINT para categorías específicas siempre debe justificarse correctamente [89].

En el lado negativo, el usuario final de OSINT podría ser un delincuente que intenta cometer un delito. Un cracker podría perfilar el objetivo para aumentar la probabilidad de éxito. Un ladrón podría analizar a los miembros de la familia para robar en casa en el mejor momento.
Un extorsionista podría publicar la información privada y personal de la víctima si no se paga un rescate.

Los desarrolladores deben tener en cuenta los aspectos antes mencionados al implementar las herramientas OSINT. En cualquier caso, por nuestro bien, las herramientas más potentes deberían estar solo al alcance de las LEA y las Agencias de Inteligencia.

G. BATALLA CONTRA EL MAL USO DE OSINT

Como ya se mencionó a lo largo de las Secciones anteriores, las potencialidades del paradigma OSINT son bastante amplias.
De hecho, sí es posible aprovechar los datos abiertos con fines de ciberseguridad y ciberdefensa, investigando así a los atacantes y/o grupos terroristas [90]. No obstante, la explotación de los datos disponibles públicamente es propensa al abuso. Es decir, los actores mal motivados pueden aprovechar la gran cantidad de información para cometer ciberagresiones, como ciberacoso, ciberchismes y cibervictimización [91]. Desafortunadamente, estos fenómenos son cada vez más frecuentes en la Red de manera alarmante, llevando a las víctimas a la angustia, la soledad, la depresión e incluso al suicidio en el peor de los casos [16]. En particular, el ciberchisme es realizado por un grupo de personas que realizan comentarios evaluativos a través de dispositivos digitales sobre alguien que no está presente. Este cibercomportamiento afecta al grupo social en el que se produce y puede dificultar las relaciones entre iguales, perjudicando a la víctima de dicho proceso [92].

En este sentido, es importante controlar que las técnicas y servicios OSINT se utilicen de forma correcta, sin lesionar los derechos y libertades de los demás [93]. Más específicamente, se podría pensar en otorgar diferentes privilegios según la categoría del usuario final, evitando así otorgar acceso completo a todo el espectro de información. Por ejemplo, los empleados pueden tener acceso a información básica para mejorar su

(p. ej., para tareas de contratación de recursos humanos), mientras que el gobierno y las fuerzas policiales pueden explorar e investigar más datos abiertos (p. ej., para cazar a un ciberdelincuente).

Finalmente, es importante señalar que OSINT está habilitando nuevas propuestas para combatir este flagelo de las ciberagresiones [94]. En este sentido, es probable que el uso indebido de OSINT se detecte correctamente con herramientas basadas en OSINT.

X. CONCLUSIÓN Y TRABAJO FUTURO

El uso generalizado de foros, redes sociales o medios de comunicación, así como la gran cantidad de datos existentes, convierten a Open Source Intelligence (OSINT) en la próxima mina de oro de Internet. La extracción de conocimiento de fuentes públicas representa una forma de resolver los problemas existentes desde una perspectiva diferente e innovadora. En concreto, la ciberseguridad y la ciberdefensa pueden verse muy beneficiadas por los resultados que este tipo de inteligencia puede ofrecer. Por lo tanto, se deben implementar procesos OSINT automatizados, capaces de llevar las investigaciones a todas partes de Internet y extender nuestra mente a través de la web.

Este documento describió el estado actual de OSINT. Reveló que la efectividad de los trabajos actuales es cuestionable debido principalmente a su pobre aplicación en escenarios reales. De hecho, faltan enfoques serios para transformar OSINT en una solución robusta y autogestionada. No obstante, sugerimos la integración de OSINT en los mecanismos de ciberdefensa existentes para pasar de los rastros técnicos atómicos de un ciberincidente al perfil del culpable o la identidad del sospechoso.

El artículo también presentó algunas técnicas OSINT para búsquedas básicas y describió las herramientas OSINT más sofisticadas en la actualidad para investigaciones avanzadas. En función de los datos disponibles y del objetivo final, una adecuada selección de la herramienta más adecuada marcaría la diferencia. Sin embargo, una combinación variada de ellos es en realidad la clave para lograr resultados plausibles.

En el contexto de España, señalamos algunos indicios que podrían confirmar que las Fuerzas y Cuerpos de Seguridad y los Servicios de Inteligencia españoles emplean OSINT en sus procedimientos internos. A pesar de ser un aspecto confidencial de su funcionamiento, OSINT es un elemento crucial en el contexto de sus investigaciones. Cabe señalar que España sería un gran territorio donde investigar, desarrollar y aplicar esta metodología debido a su madurez Open Data. De hecho, es uno de los países más transparentes de Europa, según el European Data Portal.

Como líneas de investigación futuras, el artículo describió algunos desafíos abiertos relacionados con la recopilación, el análisis y la extracción de conocimiento real de la inmersión en Internet. Aspectos como la desinformación, la privacidad y la legalidad serán destacados en el futuro de OSINT. Todavía queda un largo camino por recorrer en esta área y, para ello, la comunidad debe abordar los desafíos discutidos mediante la inclusión de técnicas avanzadas y la mejora del desempeño actual. El objetivo final de OSINT es poder garantizar el hallazgo deseado para un propósito determinado, de forma automatizada y autodirigida.

camino.

REFERENCIAS

[1] HJ Williams e I. Blum, "Definiendo inteligencia de código abierto de segunda generación (OSINT) para la empresa de defensa", RAND Corp., Santa Mónica, CA, EE. UU., Tech. Rep. RR-1964-OSD, 2018, doi: 10.7249/RR1964.

[2] M. Nouh, JR Nurse, H. Webb y M. Goldsmith, "¡Los investigadores del cibercrimen también son usuarios! Comprender los desafíos sociotécnicos que enfrenta la aplicación de la ley", en *Proc. Taller de seguridad usable de 2019,* febrero de 2019.

[3] A. Powell y C. Haynes, "Datos de redes sociales en investigaciones forenses digitales", en *Educación forense digital: un enfoque de aprendizaje experiencial,* X. Zhang y K.-KR Choo, Eds. Cham, Suiza: Springer, 2020, págs. 281–303.

[4] G. Bello-Orgaz, JJ Jung y D. Camacho, "Big data social: logros recientes y nuevos desafíos", *Inf. Fusión,* vol. 28, págs. 45 a 59, marzo de 2016.

[5] HL Larsen, JM Blanco, RP Pastor y RR Yager, Eds., *Uso de datos abiertos para detectar amenazas del crimen organizado: factores que impulsan el crimen futuro.* Cham, Suiza: Springer, 2017.

[6] M. Dawson, M. Lieble y A. Adeboje, "Inteligencia de código abierto: extracción de datos y análisis de vínculos para rastrear actividades terroristas", en *Tecnología de la información: nuevas generaciones,* vol. 558. Cham, Suiza: Springer, julio de 2018, págs. 1–11.

[7] F. Ali, FH Khan, S. Bashir y U. Ahmad, "Contrarrestar el terrorismo en las redes sociales en línea utilizando técnicas de minería web", en *Tecnologías y aplicaciones inteligentes,* IS Bajwa, F. Kamareddine y A. Costa, Eds. Singa poro: Springer, 2019, págs. 240–250.

[8] J. Jang-Jaccard y S. Nepal, "Una encuesta sobre las amenazas emergentes en ciberseguridad", *J. Comput. sist. ciencia,* vol. 80, núm. 5, págs. 973–993, agosto de 2014.

[9] F. Gómez Mármol, M. Gil Pérez, and G. Martínez Pérez, "No confío en las TIC: Retos de investigación en ciberseguridad", en *Trust Management* X, SM Habib, J. Vassileva, S. Mauw y M. Mühlhäuser, Eds. Cham, Suiza: Springer, 2016, págs. 129–136.

[10] P. Nespoli, D. Papamartzivanos, F. Gomez Marmol y G. Kambourakis, "Selección de contramedidas óptimas contra ataques cibernéticos: una encuesta completa sobre marcos de reacción", *IEEE Commun. Encuestas Tuts.,* vol. 20, núm. 2, págs. 1361–1396, 2.º trimestre, 2018.

[11] D. Quick y K.-K.-R. Choo, "Inteligencia forense digital: subconjuntos de datos e inteligencia de fuente abierta (DFINT+OSINT): una combinación oportuna y cohesiva", *Future Gener. computar Syst.,* vol. 78, págs. 558–567, enero de 2018.

[12] L. Ball, G. Ewan y N. Coull, "Socavando: ingeniería social utilizando recopilación de inteligencia de fuente abierta", en *Proc. En t. Conf. Saber Descubrimiento Inf. Retr.,* 2012, págs. 275–280.

[13] Z. Jin, J. Cao, Y. Zhang y J. Luo, "Verificación de noticias mediante la explotación de puntos de vista sociales conflictivos en microblogs", en *Proc. XIII Conferencia AAAI. Artefacto Intel. (AAAI),* 2016, págs. 2972–2978.

[14] J. Simola, "Problemas de privacidad y protección de infraestructuras críticas", en *Emerging Cyber Threats and Cognitive Vulnerabilities,* V. Benson y J. Mcalaney, Eds. Académico, 2020, págs. 197–226.

[15] M. Kandias, L. Mitrou, V. Stavrou y D. Gritzalis, "¿De qué lado estás? Un nuevo panóptico versus privacidad", en *Proc. Internacional IEEE Conf. Seguro Criptogr. (SECRYPT),* Reykjavik, Islandia, julio de 2013, págs. 1–13.

[16] LR Betts y KA Spenser, "Desarrollo de las experiencias de victimización cibernética y escalas de conductas de ciberacoso", *J. Genet. Psychol.,* vol. 178, núm. 3, págs. 147–164, mayo de 2017.

[17] J. Pastor-Galindo, P. Nespoli, FG Mármol y GM Pérez, "OSINT es la próxima mina de oro de Internet: España como territorio inexplorado", en *Proc. 5to Nat. Conf. Ciberseguridad. (JNIC),* Cáceres, España, 2019.

[18] F. Tabatabaei y D. Wells, "Osint en el contexto de la seguridad cibernética", en *Investigación de inteligencia de fuente abierta: de la estrategia a la implementación,* B. Akhgar, PS Bayerl y F. Sampson, Eds. Cham, Suiza: Springer, 2016, págs. 213–231.

[19] H. Chen, RHL Chiang y VC Storey, "Business intelligence and analytics: From big data to big impact", *MIS Quart.,* vol. 36, núm. 4, págs. 1165–1188, 2012.

[20] V. Santarcangelo, G. Oddo, M. Pilato, F. Valenti y C. Fornaro, "Social opinion mining: An approach for Italian language", en *Proc. 3° Int. Conf. Future Internet Things Cloud,* Roma, Italia, agosto de 2015, págs. 693–697.

[21] M. Kandias, D. Gritzalis, V. Stavrou y K. Nikoloulis, "Detección del nivel de estrés a través del patrón de uso de OSN y análisis de cronicidad: un módulo de inteligencia de amenazas OSINT", *Comput. Seguridad,* vol. 69, págs. 3 a 17, agosto de 2017.

[22] B. Senekal y E. Kotzé, "Inteligencia de fuente abierta (OSINT) para el monitoreo de conflictos en la Sudáfrica contemporánea: Desafíos y oportunidades en un contexto de big data", *Afr. Seguro Rev.,* vol. 28, núm. 1, págs. 19 a 37, enero de 2019.

[23] D.-Y. Kao, Y.-T. Chao, F. Tsai y C.-Y. Huang, "Analítica de evidencia digital aplicada en investigaciones de delitos cibernéticos", en *Proc. Conferencia IEEE apl., inf. Neto. Seguro (AINS),* noviembre de 2018, págs. 117–122.

[24] RP Pastor y HL Larsen, "Escaneo de datos abiertos para la detección de amenazas emergentes del crimen organizado—El proyecto ePOOLICE", en *Uso de datos abiertos para detectar amenazas del crimen organizado.* Cham, Suiza: Springer, 2017, págs. 47–71.

[25] C. Aliprandi, J. Arraiza Irujo, M. Cuadros, S. Maier, F. Melero, and M. Raffaelli, "Caper: Información, adquisición, procesamiento, explotación y reporte colaborativo para la prevención del crimen organizado, " en *HCI International 2014—Posters' Extended Abstracts,* C. Stephanidis, Ed. Cham, Suiza: Springer, 2014, págs. 147–152.

[26] T. Delavallade, P. Bertrand y V. Thouvenot, "Extracción de indicadores de delitos futuros de las redes sociales", en *Uso de datos abiertos para detectar amenazas de delincuencia organizada.* Cham, Suiza: Springer, 2017, págs. 167–198.

[27] MJ Hernández, CC Pinzón, DO Díaz, JCC García y RA Pinto, "Inteligencia de fuente abierta (OSINT) en un contexto colombiano y análisis de sentimiento", *Rev. V'inculos, Ciencia, Tecnol. Sociedad,* vol. 15, núm. 2, págs. 195–214, 2018.

[28] *Proyecto de mejora de la diversidad para la seguridad de la información y la gestión de eventos.* Consultado: 9 de enero de 2020. [En línea]. Disponible: http://disiem project.eu/

[29] S. Lee y T. Shon, "Marco de inspección de amenazas cibernéticas de base de inteligencia de código abierto para infraestructuras críticas", en *Proc. Tecnología del futuro. Conf. (FTC),* San Francisco, CA, EE. UU., diciembre de 2016, págs. 1030–1033.

[30] M. Edwards, R. Larson, B. Green, A. Rashid y A. Baron, "La búsqueda de oro: análisis automático de superficies de ataque de ingeniería social en línea", *Comput. Seguridad,* vol. 69, págs. 18 a 34, agosto de 2017.

[31] MG Lozano, J. Brynielsson, U. Franke, M. Rosell, E. Tjornhammar, S. Varga y V. Vlassov, "Evaluación de la veracidad de los datos en línea", *Decis. Sistema de apoyo,* vol. 129, febrero de 2020, art. no. 113132.

[32] BLW Wong, "Fluidez y rigor: abordaje de las consideraciones de diseño para herramientas y procesos de osint", en *Investigación de inteligencia de código abierto: de la estrategia a la implementación,* B. Akhgar, PS Bayerl y F. Sampson, Eds. Cham, Suiza: Springer, 2016, págs. 167–185.

[33] G. Kalpakis, T. Tsikrika, N. Cunningham, C. Iliou, S. Vrochidis, J. Middleton e I. Kompatsiaris, *OSINT y la Dark Web.* Cham, Suiza: Springer, 2016, págs. 111–132.

[34] MK Bergman, "Libro blanco: La web profunda: valor oculto aflorando" *J. Electron. Editorial,* vol. 7, núm. 1, agosto de 2001.

[35] M. Schafer, M. Fuchs, M. Strohmeier, M. Engel, M. Liechti y V. Lenders, "BlackWidow: Monitoreo de la web oscura para la información de seguridad cibernética", en *Proc. 11 Int. Conf. Cyber Conflict (CyCon),* Tallin, Estonia, mayo de 2019, págs. 1 a 21.

[36] A. Gandomi y M. Haider, "Más allá de la exageración: conceptos, métodos y análisis de big data", *Int. J.Inf. Administrar.,* vol. 35, núm. 2, págs. 137 a 144, abril de 2015.

[37] A. Barnea, "Grandes datos y contrainteligencia en los países occidentales", *Int. j Intel. Contrainteligencia,* vol. 32, núm. 3, págs. 433–447, julio de 2019.

[38] T. Day, H. Gibson y S. Ramwell, "Fusión de datos OSINT y no OSINT", en *Open Source Intelligence Investigation.* Cham, Suiza: Springer, 2016, págs. 133–152.

[39] CS Fleisher, "Uso de datos de fuente abierta para desarrollar inteligencia competitiva y de marketing", *Eur. J. Marketing,* vol. 42, núm. 7/8, págs. 852–866, julio de 2008.

[40] FG Marmol, MG Perez y GM Perez, "Informar sobre contenido ofensivo en las redes sociales: hacia un enfoque de evaluación basado en la reputación", *IEEE Internet Comput.,* vol. 18, núm. 2, págs. 32 a 40, marzo de 2014.

[41] S. Gong, J. Cho y C. Lee, "Un método de comparación de confiabilidad para el análisis de validez OSINT", *IEEE Trans. Ind. Informat.,* vol. 14, núm. 12, págs. 5428–5435, diciembre de 2018.

[42] M. Zago, P. Nespoli, D. Papamartzivanos, MG Perez, FG Marmol, G. Kambourakis y GM Perez, "Screening out social bots interfer ence: Are there any silver bullets?" *IEEE Commun. Mag.,* vol. 57, núm. 8, págs. 98–104, agosto de 2019.

[43] GR Weir, "Las limitaciones de la automatización de osint: comprender la pregunta, no la respuesta", en *Automatización de la inteligencia de fuente abierta,* R. Layton y PA Watters, Eds. Boston, MA, EE. UU.: Syngress, 2016, págs. 159–169.

[44] P. Casanovas, "Ciberguerra y crimen organizado. Un modelo regulatorio y meta-modelo para inteligencia de fuente abierta (OSINT)," en *Ética y Políticas para Operaciones Cibernéticas.* Cham, Suiza: Springer, 2017, págs. 139–167.

[45] H. Bean, "¿Es la inteligencia de fuentes abiertas una cuestión ética?" en *Investigación sobre problemas sociales y políticas públicas,* vol. 19, S. Maret, ed. Bingley, Reino Unido: Emerald Group Publishing Limited, 2011, págs. 385–402.

[46] B. Liu y L. Zhang, "Una encuesta sobre análisis de sentimientos y minería de opiniones", en *Mining Text Data.* Boston, MA, EE. UU.: Springer, 2012, págs. 415–463.

[47] P. Ranade, S. Mittal, A. Joshi y K. Joshi, "Uso de redes neuronales profundas para traducir inteligencia de amenazas multilingüe", en *Proc. Internacional IEEE Conf. Intel. Seguro Informar. (ISI),* noviembre de 2018, págs. 238–243.

[48] Y. Ghazi, Z. Anwar, R. Mumtaz, S. Saleem y A. Tahir, "Un enfoque basado en el aprendizaje automático supervisado para extraer automáticamente inteligencia de amenazas de alto nivel de fuentes no estructuradas", en *Proc. En t. Conf. Fronteras Inf. Tecnología (FIT),* diciembre de 2018, págs. 129–134.

[49] S. Noubours, A. Pritzkau y U. Schade, "La PNL como ingrediente esencial de los marcos OSINT efectivos", en *Proc. Mil. común información sist. Conf.,* octubre de 2013, págs. 1–7.

[50] Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan y J. Han, "Una encuesta sobre el descubrimiento de la verdad", *SIGKDD Explor. Boletín.,* vol. 17, núm. 2, págs. 1 a 16, febrero de 2016.

[51] T. Vopham, JE Hart, F. Laden y YY Chiang, "Tendencias emergentes en inteligencia artificial geoespacial (geoAI): aplicaciones potenciales para la epidemiología ambiental", *Environ. Salud,* vol. 17, núm. 1 de abril de 2018.

[52] S. Stieglitz, M. Mirbabaie, B. Ross y C. Neuberger, "Análisis de redes sociales: desafíos en el descubrimiento de temas, recopilación y preparación de datos", *Int. J.Inf. Administrar.,* vol. 39, págs. 156–168, abril de 2018.

[53] L. Serrano, M. Bouzid, T. Charnois, S. Brunessaux y B. Grilheres, "Extracción y agregación de eventos para inteligencia de fuente abierta: del texto al conocimiento", *Proc. En t. Conf. Herramientas Artif. Intel. (ICTAI),* 2013, págs. 518–523.

[54] N. Kim, S. Lee, H. Cho, B.-I. Kim y M. Jun, "Diseño de un sistema de recopilación de información sobre ciberamenazas para la correlación de ciberataques", en *Proc. En t. Conf. Tecnología de la plataforma. Service (PlatCon),* enero de 2018, págs. 1 a 6.

[55] S. Pournouri, S. Zargari y B. Akhgar, "Una investigación sobre el uso de técnicas de clasificación en la predicción del tipo de objetivos en ataques cibernéticos", en *Proc. IEEE 12 Int. Conf. Global Secur., Saf. Sostenibilidad (ICGS3),* enero de 2019, págs. 202–212.

[56] I. Deliu, C. Leichter y K. Franke, "Extracción de inteligencia sobre amenazas cibernéticas de los foros de hackers: máquinas de vectores de soporte versus redes neuronales convolucionales", en *Proc. Internacional IEEE Conf. Big Data (Big Data),* diciembre de 2017, págs. 3648–3656.

[57] G. de la Torre-Abaitua, LF Lago-Fernández y D. Arroyo, "Un marco basado en compresión para la detección de anomalías en fuentes de datos heterogéneas", 2019, *arXiv:1908.00417.* [En línea]. Disponible: https://arxiv.org/abs/1908.00417 [58] R. Azevedo, I. Medeiros y A. Bessani, "PURE: Generating quality Threat Intelligence by Clustering and Corlating OSINT", en *Proc. 18º IEEE Internacional. Conf. Confianza, seguridad. Cómputo de privacidad. Comun./13th IEEE Int. Conf. Ciencia de los grandes datos Ing. (TrustCom/BigDataSE),* agosto de 2019, págs. 483–490.

[59] M.-H. Wang, M.-H. Tsai, W.-C. Yang y C.-L. Lei, "Categorización de infecciones mediante un codificador automático profundo", en *Proc. Conferencia IEEE computar común Talleres (INFOCOM WKSHPS),* abril de 2018, págs. 1 y 2.

[60] H. Pellet, S. Shiaeles y S. Stavrou, "Localización de usuarios de redes sociales y perfiles de su movimiento", *Comput. Segur.,* vol. 81, págs. 49–57, marzo de 2019.

[61] R. Wang, W. Ji, M. Liu, X. Wang, J. Weng, S. Deng, S. Gao y C.-A. Yuan, "Revisión de datos de extracción de múltiples fuentes de datos", *Reconocimiento de patrones. Lett.,* vol. 109, págs. 120–128, julio de 2018.

[62] R. Layton, C. Perez, B. Birregah, P. Watters y M. Lemercier, "Enlace de información indirecta para OSINT a través del análisis de autoría de alias", en *Trends and Applications in Knowledge Discovery and Data Mining,* J. Li, L. Cao, C. Wang, KC Tan, B. Liu, J. Pei y VS Tseng, Eds. Berlín, Alemania: Springer, 2013, págs. 36–46.

[63] A. Chaabane, P. Manils y MA Kaafar, "Excavando en el tráfico anónimo: un análisis profundo de la red anonimizadora de Tor", en *Proc. 4to Int. Conf. Neto. sist. Seguro (NSS),* septiembre de 2010, págs. 167–174.

[64] S. Pastrana, A. Hutchings, A. Caines y P. Buttery, "Characteriz ing eve: Analyzing cybercrimeactors in a large underground forum", en *Research in Attacks, Intrusions, and Defenses,* M. Bailey , T. Holz, M. Stamatogiannakis y S. Ioannidis, eds. Cham, Suiza: Springer, 2018, págs. 207–227.

[65] W. Tounsi y H. Rais, "Una encuesta sobre inteligencia técnica de amenazas en la era de los ciberataques sofisticados", *Comput. Segur.,* vol. 72, págs. 212 a 233, enero de 2018.

[66] C. Sauerwein, I. Pekaric, M. Felderer y R. Breu, "Un análisis y clasificación de las fuentes de datos de seguridad de la información pública utilizadas en la investigación y la práctica", *Comput. Segur.,* vol. 82, págs. 140–155, mayo de 2019.

[67] R. Layton, "Atribución relativa del ciberataque", en *Automatización de la inteligencia de fuente abierta: Algoritmos para OSINT,* R. Layton y PA Watters, Eds. Boston, MA, EE. UU.: Syngress, 2016, págs. 37–60.

[68] V. Mavroeidis y S. Bromander, "Modelo de inteligencia de amenazas cibernéticas: una evaluación de taxonomías, estándares de intercambio y ontologías dentro de la inteligencia de amenazas cibernéticas", en *Proc. EUR. Intel. Seguro Informar. Conf. (EISIC),* Atenas, Grecia, septiembre de 2017, págs. 91–98.

[69] S. Bromander, A. Jøsang y M. Eian, "Modelado semántico de ciberamenazas", en *Proc. XI Conferencia Tecnología semántica. Intell., Defense, Secur.,* Fairfax, VA, EE. UU., noviembre de 2016, págs. 74–78.

[70] O. Akinrolabu, I. Agrafiotis y A. Erola, "El desafío de detectar ataques sofisticados: conocimientos de los analistas del SOC", en *Proc. 13 Int. Conf. Disponibilidad, rel. Seguro (ARES),* 2018, págs. 55:1–55:9.

[71] D. Lande y E. Shnurko-Tabakova, "OSINT como parte del sistema de ciberdefensa", *Theor. aplicación Ciberseguridad,* vol. 1, no. 1, 2019.

[72] B. Akhgar, "Osint como parte integral del aparato de seguridad nacional", en *Investigación de inteligencia de fuente abierta: de la estrategia a la implementación,* B. Akhgar, PS Bayerl y F. Sampson, Eds. Cham, Suiza: Springer, 2016, págs. 3–9.

[73] J. Chae, D. Graham, A. Henderson, M. Matthews, J. Orcutt y MS Song, "Un enfoque sistémico para evaluar las herramientas de inteligencia de código abierto actuales y emergentes del ejército", en *Proc. Internacional IEEE sist. Conf. (SysCon),* abril de 2019, págs. 1 a 5.

[74] D. Trottier, "Inteligencia de fuente abierta, redes sociales y aplicación de la ley: visiones, limitaciones y críticas", *Eur. J. Cultural Stud.,* vol. 18, nn. 4–5, págs. 530–547, agosto de 2015.

[75] L. Rocher, JM Hendrickx y Y.-A. de Montjoye, "Estimación del éxito de las reidentificaciones en conjuntos de datos incompletos utilizando modelos generativos", *Nature Commun.,* vol. 10, núm. 1, pág. 3069, 2019.

[76] RS Portnoff, S. Afroz, G. Durrett, JK Kummerfeld, T. Berg-Kirkpatrick, D. Mccoy, K. Levchenko y V. Paxson, "Herramientas para el análisis automatizado de mercados cibercriminales", en *Proc . 26 Int. Conf. World Wide Web (WWW),* 2017, págs. 657–666.

[77] E. Ferrara, P. De Meo, G. Fiumara y R. Baumgartner, "Extracción de datos web, aplicaciones y técnicas: una encuesta", *Knowl.-Based Syst.,* vol. 70, págs. 301 a 323, noviembre de 2014.

[78] IH Witten, E. Frank, MA Hall y CJ Pal, *Minería de datos: herramientas y técnicas prácticas de aprendizaje automático.* San Mateo, CA, EE. UU.: Morgan Kaufmann, 2017.

[79] A. Caliskan, JJ Bryson y A. Narayanan, "La semántica derivada automáticamente de los corpus lingüísticos contiene sesgos similares a los humanos", *Science,* vol. 356, núm. 6334, págs. 183–186, abril de 2017.

[80] C. Eldridge, C. Hobbs y M. Moran, "Fusionar algoritmos y analistas: inteligencia de fuente abierta en la era de los 'grandes datos'", *Intell. Nat. Segur.,* vol. 33, núm. 3, págs. 391–406, abril de 2018.

[81] X. Yin, J. Han y P. Yu, "Descubrimiento de la verdad con múltiples proveedores de información en conflicto en la Web", *IEEE Trans. Saber Ing. de datos,* vol. 20, núm. 6, págs. 796–808, junio de 2008.

[82] AS Hulnick, "El dilema de la inteligencia de fuentes abiertas: ¿OSINT es realmente inteligencia?" en *The Oxford Handbook of National Security Intelligence,* LK Johnson, Ed. Oxford, Reino Unido: Universidad de Oxford. Prensa, septiembre de 2010.

[83] K. Shu, A. Sliva, S. Wang, J. Tang y H. Liu, "Detección de noticias falsas en las redes sociales: una perspectiva de minería de datos", *SIGKDD Explor. Newsl.,* vol. 19, núm. 1, págs. 22 a 36, septiembre de 2017.

[84] BH Miller, "Inteligencia de fuente abierta (OSINT): ¿Un oxímoron?" *Ent. j Intel. Contrainteligencia,* vol. 31, núm. 4, págs. 702–719, octubre de 2018.

[85] G. Suarez-Tangil, M. Edwards, C. Peersman, G. Stringhini, A. Rashid y M. Whitty, "Desmantelamiento automático del fraude de citas en línea", *IEEE Trans. información Seguridad forense,* vol. 15, págs. 1128–1137, 2020.

[86] J. Rajamäki y J. Simola, "¿Cómo aplicar la privacidad por diseño en osint y análisis de big data?" en *Proc. EUR. Conf. información Seguridad de guerra. (ECCWS),* julio de 2019, págs. 364–371.

[87] JH Hoepman, "Estrategias de diseño de privacidad", en *Proc. IFIP Adv. información Comun. Tecnología,* vol. 428, 2014, págs. 446–459.

[88] G. Hribar, I. Podbregar y T. Ivanuša, "OSINT: ¿Una zona gris?" *Int. j Intel. Contrainteligencia,* vol. 27, núm. 3, págs. 529–549, 2014.

[89] Q. Eijkman y D. Weggemans, "Dilemas de privacidad e inteligencia de fuente abierta: ¿Es hora de reevaluar la responsabilidad estatal?" *Secur. Tararear. Derechos,* vol. 23, núm. 4, págs. 285 y 296, 2013.

[90] P. Mitzias, I. Kompatsiaris, E. Kontopoulos, J. Staite, T. Day, G. Kalpakis, T. Tsikrika, H. Gibson, S. Vrochidis y B. Akhgar, "Desplegar tecnologías web semánticas para la fusión de información de contenido relacionado con el terrorismo y detección de amenazas en la Web", en *Proc. Internacional IEEE/WIC/ACM Conf. Inteligencia web. (WI) Companion,* 2019, págs. 193–199.

[91] GW Giumetti y RM Kowalski, "El ciberacoso es importante: examen del impacto incremental del ciberacoso en los resultados por encima del acoso tradicional en América del Norte", en *Ciberacoso en todo el mundo.* Cham, Suiza: Springer, 2016, págs. 117–130.

[92] EM Romera, M. Herrera-López, JA Casas, RO Ruiz, and R. Del Rey, "¿Cuánto ciberchisme tienen los adolescentes? Desarrollo y validación de escalas en España y Colombia", *Frontiers Psychol.,* vol. 9, págs. 1 a 10, febrero de 2018.

[93] L. Benes, "OSINT, nuevas tecnologías, educación: Ampliando oportunidades y amenazas. Un nuevo paradigma", *J. Strategic Secur.,* vol. 6, núm. 3, págs. 22 a 37, septiembre de 2013.

[94] A. López-Martínez, JA García-Díaz, R. Valencia-García, and A. Ruiz-Martínez, "CyberDect. Un enfoque novedoso para la detección del ciberacoso en Twitter", en *Tecnologías e Innovación.* Cham, Suiza: Springer, 2019, págs. 109–121.

JAVIER PASTOR-GALINDO recibió el B.Sc. y M.Sc. Licenciado en Informática por la Universidad de Murcia, España. En 2019, obtuvo un Contrato Predoctoral FPU del Ministerio de Ciencia, Innovación y Universidades de España, para desarrollar su doctorado. con el Departamento de Ingeniería de la Información y las Comunicaciones de la Universidad de Murcia. Sus intereses de investigación se centran en la inteligencia de código abierto (OSINT), la seguridad y la privacidad.

PANTALONE NESPOLI recibió el B.Sc. y maestrías en ingeniería informática de la Universidad de Napoli Federico II, Italia. Actualmente está cursando el Ph.D. Doctorado en la Universidad de Murcia, España. Sus intereses de investigación incluyen la seguridad de los sistemas de información y comunicación; más concretamente, seguridad de redes, sistemas de detección y respuesta a intrusos, y gestión de eventos e información de seguridad.

FÉLIX GÓMEZ MÁRMOL recibió el M.Sc. y doctorado Licenciado en Ingeniería Informática por la Universidad de Murcia, España. Actualmente es Investigador del Departamento de Ingeniería de la Información y las Comunicaciones de la Universidad de Murcia. Sus intereses de investigación incluyen ciberseguridad, Internet de las cosas, aprendizaje automático y algoritmos bioinspirados.

GREGORIO MARTÍNEZ PÉREZ es actualmente Profesor Titular del Departamento de Ingeniería de la Información y las Comunicaciones de la Universidad de Murcia, España. Su actividad científica se dedica principalmente a la ciberseguridad, la privacidad y las redes. Está trabajando en diferentes proyectos de investigación IST nacionales y europeos (25 en la última década) sobre estos temas, siendo el investigador principal en la mayoría de ellos. Ha publicado más de 160 artículos en actas de congresos, revistas y diarios nacionales e interna