Universidad Internacional de La Rioja (UNIR)
Máster universitario en Seguridad Informática

OSINT. Investigación, análisis y propuesta para su uso en instituciones educativas

Trabajo Fin de Máster

presentado por: Angulo Vidal, Alcides Augusto

Director/a: Rodríguez Gómez, Rafael Alejandro

Ciudad: Málaga
Fecha: Julio 2020

Resumen

La transformación digital es un hecho en todos los sectores económicos y el ámbito educativo no es ajeno a este auge. El mundo digital trae consigo más información expuesta y nuevas amenazas. El proceso por el cual se recolecta y analiza la información disponible en fuentes abiertas, para producir inteligencia, se conoce como *Open Source Intelligence* (OSINT) y su uso está muy extendido. Tras realizar un estudio del estado del arte, se indica que es fundamental que las instituciones educativas dispongan de una herramienta OSINT que permita auditar su información expuesta. En este trabajo se ha realizado un experimento para determinar qué herramientas de software libre presentan mejores resultados. Después de su análisis, se concluye para qué propósito es mejor cada una de las herramientas, y así proporcionar a estas instituciones de las herramientas con las que pueden protegerse mejor y a las Administraciones la posibilidad de extender estas auditorías a sus centros de forma periódica.

Palabras Clave: Inteligencia en fuentes abiertas, OSINT, Institución educativa, Software libre, Auditoría.

Abstract

Digital transformation is a fact in all economic sectors and the educational field is not unaware to this boom. The digital world brings with it more exposed information and new threats. The process by which information available in open sources is collected and analysed, to produce intelligence, is known as Open Source Intelligence (OSINT) and its use is widespread. After a study of the state of art, it is indicated that it is fundamental that educational institutions have an OSINT tool that allows them to audit their exposed information. In this work, an experiment has been carried out to determinate which free software tools present better results. After their analysis, it is concluded for which purpose each of the tools is best, and thus provide these institutions with the tools with which they can better protect themselves and the Administrations with the possibility of extending these audits to their centres on a regular basis.

Keywords: Open Source Intelligence, OSINT, Educational Sector, Opensource software, Audit.

Índice

Índice de tablas y figuras

1. Introducción

1.1 Motivación

Se estima que hay 4.500 millones de personas con acceso a Internet, comunicándose e intercambiando información en todo el mundo (Miniwatts Marketing Group, 2020). A esto, además hay que sumarle todo tipo de dispositivos que también están conectados. La cantidad de datos públicos que se generan y que están disponibles es enorme y su tendencia es que siga creciendo. Supone un gran reto para nuestra sociedad porque constantemente aparecen nuevos riesgos y oportunidades. Esto es así ya que a partir de estos datos se puede obtener información de gran valor y utilidad mediante técnicas OSINT.

Open Source Intelligence (OSINT) se define como “inteligencia que se produce a partir de información pública y disponible que es recolectada, explotada y diseminada de manera oportuna a un público apropiado con el fin de abordar un requisito de inteligencia específica”. (Williams & Blum, 2018). Hoy en día hace referencia a un conjunto de técnicas de recolección de información de diferentes fuentes abiertas, especialmente en Internet. Estas fuentes abiertas pueden ser, por ejemplo: medios de comunicación en masa, redes sociales, foros, blogs, conferencias, datos públicos gubernamentales, datos comerciales o de empresas.

Su origen es militar, durante la segunda guerra mundial Estados Unidos crea el Servicio de Información de Radiodifusión Extranjera para recopilar información pública de otros países. Actualmente los gobiernos, servicios de inteligencia y Fuerzas y Cuerpos de Seguridad de los diferentes estados usan OSINT para combatir el crimen. Pero no se utiliza solo para asuntos de Estado, sino que también se usa en todo tipo de sectores con diferentes finalidades. Multinacionales, bancos y todo tipo de industrias acuden a OSINT porque saben que, “los datos de fuentes abiertas juegan un rol significante en nuestra sociedad actual y más aún con la transformación digital hacia un mundo hiperconectado. Se ha convertido en una parte importante en las actividades de las empresas, para el análisis de negocio, impacto y planes estratégicos” (Cerny et al., 2019).

Es muy amplio el tipo de organizaciones y sectores que utilizan OSINT. Se puede utilizar para conocer la reputación online de una persona o empresa, para realizar estudios sociológicos, psicológicos, etc. También para realizar auditorías de organismos de todo tipo con el fin evaluar su seguridad y privacidad. Por otro lado, estas técnicas también son usadas por los cibercriminales. Para llevar estas técnicas a cabo hay que seguir una metodología. OSINT,

sin embargo, todavía necesita una metodología clara (Williams & Blum, 2018). Y Durante este proceso hay que hacer uso de herramientas de todo tipo dependiendo del objetivo de la investigación.

En esta era de información y también de desinformación, el sector de las instituciones educativas está inmerso en los procesos de digitalización y de transformación, impulsados por la tecnología y por los nuevos modelos de hacer las cosas en un mundo cada vez más digital. Existe un mayor riesgo de que la información quede expuesta y en estas instituciones estos datos son de especial relevancia. La motivación de este trabajo es proporcionar una herramienta de análisis OSINT enfocado a dicho sector. Una herramienta de software libre que, en base a su sencillez de uso, por el tipo de datos que genera una institución de este tipo y la estructura organizativa de estos centros, ofrezca a estas instituciones mecanismos para evaluar su nivel de exposición de datos. Finalmente se propondrán una serie de recomendaciones en base a la información que se han podido detectar con las herramientas.

1.2 Planteamiento del trabajo

Para proporcionar una solución OSINT a las instituciones educativas, lo primero de todo es conocer el estado del arte de esta técnica. Estudiarlo, conocer sus metodologías, tendencias, las aplicaciones que tiene y las herramientas más destacadas. Esto va permitir conocer los diferentes enfoques que se le puede dar y también saber qué sectores pueden tener similitudes en sus objetivos con las instituciones educativas. Estudiar el nivel de exposición en Internet de este tipo de instituciones, su evolución y necesidades servirá para dar forma a la propuesta. Para llevarlo a cabo se revisarán las herramientas apoyándonos en sus manuales, se consultarán las diferentes metodologías, estudios y casos de uso en publicaciones científicas y libros relevantes. Una vez adquirido todo este conocimiento, se plantea un experimento OSINT aplicado a un objetivo del sector educativo. Con los resultados obtenidos, y tras un análisis en profundidad se sacan las conclusiones para proponer la herramienta que cumpla con los objetivos.

1.3 Estructura

La estructura del trabajo es la siguiente:

- Capítulo 1. Introducción . En este capítulo se presenta la motivación de la realización de este trabajo y el planteamiento con el que se pretende abordar.

- Capítulo 2. Estado del arte. Se presenta el contexto y estado del arte donde se estudiará en profundidad en qué consiste OSINT y sus principales herramientas. Se presentan trabajos actuales relacionados con el objetivo y las partes interesadas. Para finalizar se ve el estado actual en el sector educativo en el mundo digital y sus particularidades respecto a OSINT.

- Capítulo 3. Objetivos concretos y metodología. Se presentarán los objetivos y la metodología con el que se va abordar el estudio. Se desglosan los principales objetivos que se busca cumplir en la realización del presente trabajo fin de máster.

- Capítulo 4. Descripción del experimento. Se describe detalladamente cómo se van a realizar las pruebas de las herramientas OSINT a un objetivo del ámbito educativo. Se explica cuál es el entorno de trabajo, motivo de elección del objetivo y que pruebas se van a realizar con cada una de las herramientas.

- Capítulo 5. Presentación de los resultados. Se presenta una descripción de los resultados obtenidos con cada una las herramientas. Se detallan los datos relevantes y principales observaciones.

- Capítulo 6. Discusión de los resultados. Se hace un análisis exhaustivo de los resultados, por ende, una valoración de las herramientas en base a nuestro objetivo, lo que determina una propuesta de uso y finaliza con una pequeña guía de buenas prácticas a seguir por los centros educativos.

- Capítulo 7. Conclusiones y trabajo futuro. En este último capítulo se presentan las conclusiones y líneas de trabajo futuro que pueden ser desarrolladas a partir de esta investigación.

2. Estado del arte

En este capítulo se presentan los pilares sobre el cual se sustentará este trabajo. Se estructura en las siguientes partes. Primero se estudiará OSINT para saber qué es, cómo funciona, qué problemas tiene y cuáles son sus ventajas. Como segundo apartado se hará un estudio exhaustivo de las herramientas principales, qué peculiaridades presentan, cuáles son sus puntos fuertes y debilidades. El tercer punto es presentar que trabajos relativos a OSINT existen actualmente que se pueden parecer o que se pueden aplicar al ámbito de este trabajo. El cuarto punto, servirá para conocer el estado actual de las instituciones educativas en su digitalización y de esta manera en el último punto perfilar y justificar lo que hace diferente este ámbito al resto y de la necesidad de aportarle una solución OSINT.

2.1 OSINT

De acuerdo con (*Nato OSINT Handbook v1.2 - jan 2002*, 2001) hay cuatro categorías de información abierta e inteligencia: *Open Source Data* (OSD); *Open Source Information* (OSIF); *Open Source Intelligence* (OSINT) y *Validated* OSINT (OSINT-V). A OSINT lo define como: Información que ha sido deliberadamente descubierta, discriminada, destilada y difundida a una audiencia, generalmente al comandante y su personal inmediato, con el fin de abordar una pregunta específica. OSINT en otras palabras aplica el proceso probado de inteligencia a la amplia diversidad de fuentes abiertas de información, y crea inteligencia.

Las fuentes de OSINT se diferencian de otros tipos inteligencia porque estas deben ser accesibles legalmente por el público sin romper ninguna brecha, derechos de autor o leyes de privacidad. Es por ello que se le considera disponible públicamente. Esta distinción hace que la capacidad de recopilar fuentes OSINT sea más accesible y no esté ligado solo a la seguridad nacional. Por ejemplo, las empresas se pueden beneficiar para explotar estas fuentes y obtener inteligencia respecto a sus competidores (Fleisher, 2008).

2.1.1 Información en fuentes abiertas

La información de fuentes abiertas accesibles para OSINT se pueden encontrar de diversas formas, tanto *online* como *offline*. Por ejemplo pueden adquirirse de los siguientes sitios (Hassan & Hijazi, 2018):

- Internet, el abanico es muy amplio: foros, blogs, sitios de redes sociales, sitios para compartir videos como *YouTube.com*, wikis, registros *Whois* de nombres de dominio registrados, metadatos y archivos digitales, recursos *Dark Web*, datos de geolocalización, direcciones IP, motores de búsqueda de personas y todo lo que se puede encontrar en línea.
- Medios de comunicación tradicionales (p. ej., Televisión, radio, periódicos, libros, revistas).
- Revistas especializadas, publicaciones académicas, disertaciones, actas de congresos, perfiles de la compañía, informes anuales, noticias de la compañía, perfiles de empleados y currículums.
- Fotos y videos incluyendo metadatos.
- Información geoespacial (p. ej., Mapas y productos de imágenes comerciales).

En el mundo *online* la cantidad de información que se puede encontrar es enorme. En la Figura 1 compartida por (*bellincat*, s. f.) nos podemos hacer una pequeña idea de todas las posibilidades de fuentes abiertas que existen para realizar una investigación OSINT. En su mayoría son plataformas, herramientas y aplicaciones que proporcionan datos públicos. Se dividen en plataformas de redes sociales, Blogs, información geoespacial, herramientas forenses de imágenes y videos, movimientos aéreos y marítimos, *webcams*, etc.

Figura 1: Ilustración de posibles fuentes OSINT

Fuente: (*bellincat*, s. f.)

En la era de la información en la que estamos inmersos, las empresas, universidades y otros proveedores de fuentes OSINT están cambiando su proceso de negocio a formatos digitales. La cantidad de usuarios en redes sociales continuamente se van incrementando y el número de dispositivos de *Internet of Things* (*IoT*) serán de uso intensivo. Todo esto propiciará un inmenso incremento de datos digitales. Es decir, en el futuro la mayoría de las fuentes OSINT serán fuentes *online* (Hassan & Hijazi, 2018).

2.1.2 Ciclo de trabajo OSINT

El ciclo típico de recopilación de inteligencia comienza por identificar la necesidad de una visión adicional, seguido de la planificación de la actividad y las posibles fuentes de información. El proceso real sigue el patrón de (Fogelman-Soulié, 2008) representado en la Figura 2:

1. Recopilación: recuperación de información
2. Proceso: extracción de información
3. Analizar: análisis de tendencias / análisis de enlaces
4. Visualizar: visualización de datos
5. Colaboración

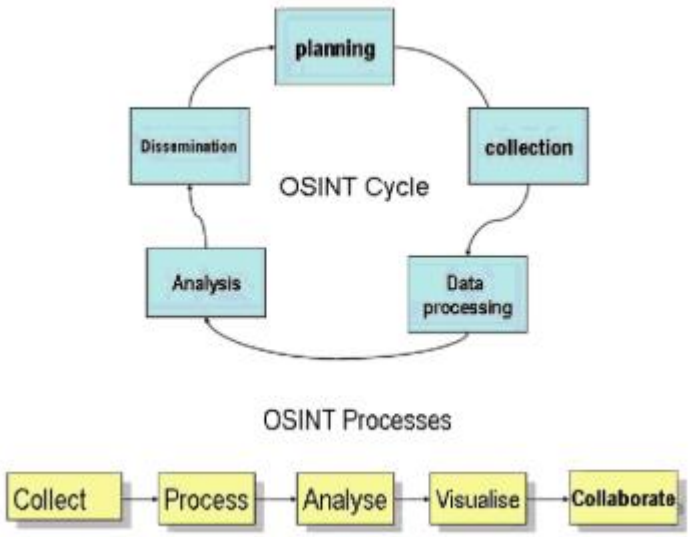


Figura 2: Ciclo OSINT

Fuente: (Fogelman-Soulié, 2008)

Según (Williams & Blum, 2018), OSINT todavía necesita una metodología clara. Por un lado, está la CIA que describe este proceso como planificación y dirección, recopilación, procesamiento, análisis y producción, y diseminación. Sin embargo por otro lado (Johnson, 2007) describe estas etapas como recopilación, procesamiento, análisis y producción, clasificación y difusión. En el contexto actual de OSINT, como podemos observar en la Figura 3, el autor se centra en cuatro pasos clave: recopilación, procesamiento, explotación y producción.

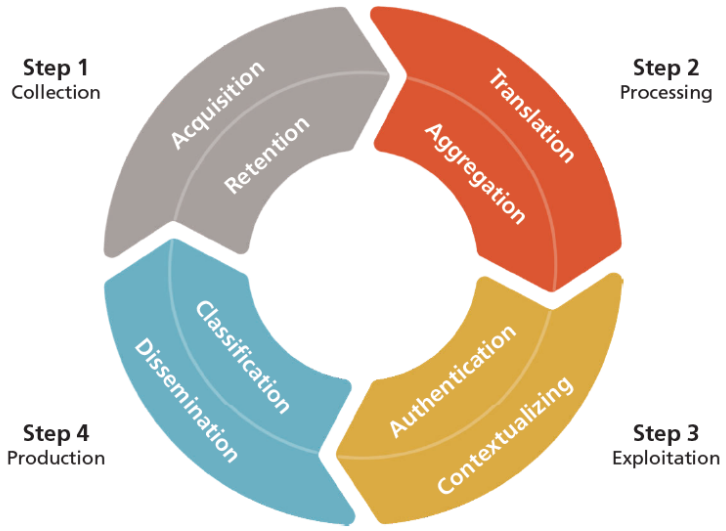


Figura 3: Ciclo OSINT

Fuente: (Williams & Blum, 2018)

1. Recopilación: Identificar la información útil. Requiere una orientación para priorizar los esfuerzos.
2. Procesamiento: Implica validar la información y hacerla usable. Actualmente la información disponible está menos estructurada por lo que este procesamiento es más complicado.
3. Explotación: A veces se le conoce análisis. Busca determinar si la información es lo que pretende ser y cuál es su valor para la inteligencia. Separar inteligencia confiable y buena de la mala.
4. Producción: Fase final en que la información se proporciona al consumidor de forma que la puede entender y utilizar.

El instituto nacional de ciberseguridad (*OSINT - La información es poder*, 2014) describe seis fases como se muestra en la Figura 4: Requisitos, identificar fuentes, Adquisición, procesamiento, análisis y presentación de inteligencia.

Figura 4: Fases OSINT. INCIBE

Fuente: (*OSINT - La información es poder*, 2014)

Respecto al modelo de fases del modelo anterior, en realidad solo añadimos una, que es la fase de requisitos, ya que las fases de identificar fuentes de información y la adquisición viene englobadas en la fase de recopilación.

- Requisitos: Condiciones que se deben satisfacer para conseguir el objetivo de resolver el problema.
- Identificar fuentes de información: A partir de los requisitos especificar las fuentes y concretar las más relevantes para optimizar el proceso.
- Adquisición: Recopilar los datos.

Por las fuentes consultadas podemos decir que las fases están definidas y podemos guiarnos por este proceso a la hora de trabajar con OSINT. (Pastor-Galindo et al., 2020) se centra en tres fases para explicar su modelo: Recopilación, análisis y extracción de conocimiento. En la Figura 5 propone un marco para llevar a cabo investigaciones OSINT. Con su esquema pretende optimizar el análisis de los resultados realizada en la recopilación y maximizar la extracción de conocimiento.

En su enfoque, una vez realizado el análisis de los datos recopilados, la salida de ese análisis lo sintetiza en tres grupos: Información personal, Información organizativa e información de red. Para finalizar con la fase de extracción de inteligencia mediante técnicas avanzadas de minado de datos e inteligencia artificial.

El autor también menciona que en esta segunda y tercera fase se hacen uso de tecnologías de procesamiento de datos. Hay que tener en cuenta que estas aplicaciones recopilan mucha información de fuentes de datos predefinidas, pero en OSINT las fuentes no son así.

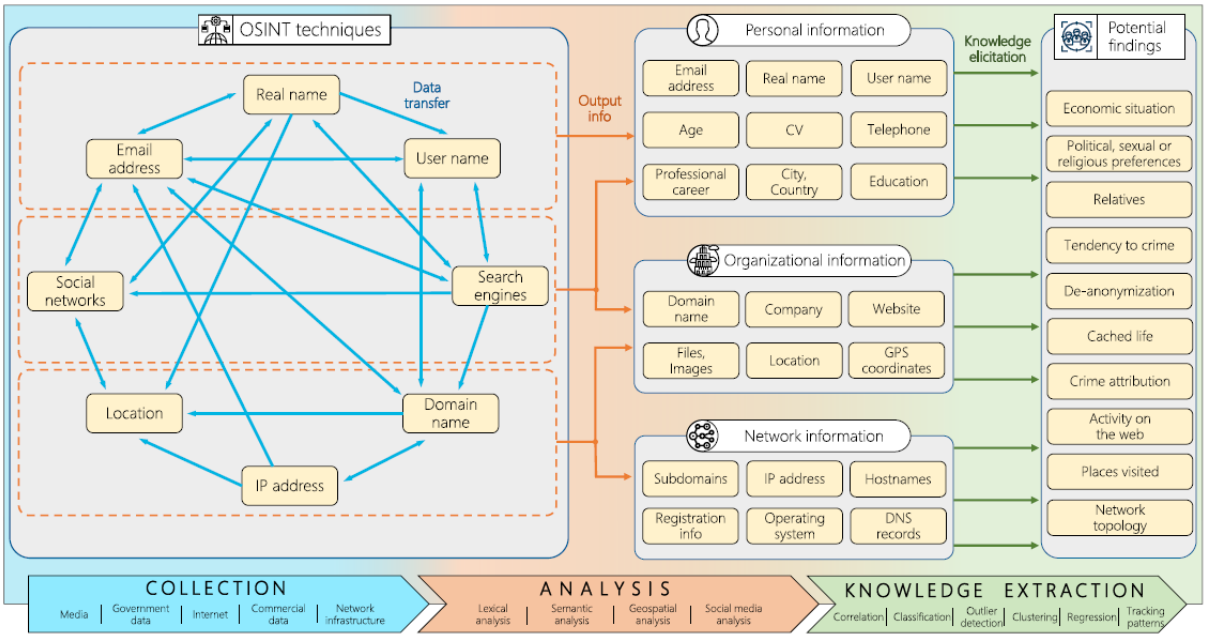


Figura 5: Principales flujos de trabajo OSINT

Fuente: (Pastor-Galindo et al., 2020)

2.1.3 Ventajas y retos de OSINT

En la era de la información no se puede subestimar el rol que juega OSINT en las diferentes áreas de la inteligencia. De manera general estas son algunas de las principales ventajas y retos según Hassan & Hijazi, (2018):

- Menor riesgo: Las otras formas de inteligencia requieren de personas sobre el terreno, recopilar información de fuentes abiertas en comparación no supone ningún riesgo.

- Rentable: Se puede usar OSINT con recursos muy limitados. Otros tipos de inteligencia por ejemplo pueden requerir mayor infraestructura y personal.
- Facilidad de accesibilidad: las fuentes de OSINT siempre están disponibles, sin importar dónde se encuentre, y siempre están actualizadas.
- Problemas legales: Al estar publicada toda esta información es menos probable que se esté infringiendo alguna licencia o derecho de autor.
- Ayuda a investigaciones financieras: Permite a los gobiernos detectar a los evasores de impuestos. A gente conocida se le ha monitorizado su estilo de vida a través de sus redes sociales facilitando a los inspectores detectar posibles fraudes.
- Lucha contra la falsificación en línea: las técnicas OSINT se pueden utilizar para encontrar productos / servicios falsos y para hacer cumplir la ley para cerrar dichos sitios o enviar advertencias a los usuarios para que dejen de tratar con ellos.
- Mantener la seguridad nacional y la estabilidad política: Identificación y prevención en el ámbito militar o de la seguridad nacional.

Actualmente son muchos los beneficios de usar inteligencia de fuentes abiertas, pero tampoco debe dejarse de lado los desafíos que conllevan. Esta disciplina como cualquier otra disciplina en inteligencia implica también una dualidad. Va a suponer, por un lado, beneficios y desafíos que generalmente son los mismos, pero visto desde un prisma diferente. Los principales desafíos son:

- El volumen de datos disponibles, que en principio es una gran ventaja. Se puede convertir en un gran problema el poder discernir información relevante.
- La fiabilidad de las fuentes, En el contexto de la inteligencia, deben ser verificadas exhaustivamente antes de poder confiar en ellas.
- Los esfuerzos humanos, gran cantidad de datos suponen un esfuerzo en la recolección, se necesita saber si son datos fiables, por lo que necesitan ser contrastados, para asegurar si relevancia y fiabilidad. Esto consume tiempo y recursos humanos.

Tabla 1: Resumen pros y contras OSINT

Pros	Contras
Gran cantidad de información disponible	Complejidad de la gestión de datos
Alta capacidad de cómputo	Información no estructurada
Big data y *machine learnin*g	Desinformación
Tipos de datos complementarios	Fiabilidad de las fuentes de datos
Propósito flexible y amplio alcance	Fuertes consideraciones éticas y legales

Fuente: (Pastor-Galindo et al., 2020)

En la Tabla 1 se puede observar de manera resumida, estos pros y contras desde el prisma de los desarrolladores e ingenieros. Este punto de vista técnico expone una serie de beneficios, pero también algunas dificultades, que se detallan a continuación. (Pastor-Galindo et al., 2020)

- Gran cantidad de información disponible. Como se ha mencionado anteriormente la cantidad de datos de fuentes abiertas es enorme. Redes sociales, documentos e informes del gobierno, contenido multimedia, periódicos e inclusive la *Deep Web* y *Dark Web*.
- Alta capacidad de cómputo. Los avances tecnológicos han facilitado que operaciones de recopilación, procesamiento, análisis y almacenamiento se apliquen a gran cantidad de datos de diferentes tipos de fuentes abiertas.
- *Big data* y *machine learning*. Técnicas de minería de datos y algoritmos de aprendizaje automático han facilitado la toma de decisiones y facilitado la detección de correlaciones que se escapan al humano.
- Tipos de datos complementarios. Posibilidad de alimentar a OSINT con otros tipos de información. Este hecho significa que OSINT puede ser aún más efectivo si podemos agregar información externa y complementar investigaciones. Por ejemplo, se podría combinar OSINT con ingeniería social para perfilar un objetivo.
- Propósito flexible y amplio alcance. Las investigaciones pueden extenderse a muchos problemas y puede recopilar información todo sobre el ciberespacio. Puede usarse para fines económicos, psicológica, estratégica, periodística, laboral o de seguridad aspectos, entre otros.

Sin embargo, el uso de fuentes abiertas presenta las siguientes desventajas:

- Complejidad de la gestión de datos. Como hemos comentado antes, hay que tener técnicas avanzadas y recursos significativos para garantizar una recolección, procesamiento y análisis de alta calidad.
- Información no estructurada. La información pública disponible en Internet está desorganizada. Los datos recopilados por OSINT son tan heterogéneos que resulta difícil de clasificar, vincular y examinar dichos datos para extraer relaciones relevantes y conocimiento. OSINT requiere mecanismos para homogeneizar la información no estructurada para poder explotarla.
- Desinformación. Redes sociales y medios de comunicación están llenos de opiniones subjetivas, noticias falsas y bulos. Esta información inexacta debe ser tenida en cuenta mediante la implementación de mecanismos. En OSINT de sebe tratar siempre con información que se pueda confiar.

- Fiabilidad de las fuentes de datos. Los datos recopilados deben provenir de fuentes revisadas y confiables. Pero también hay que tratar con las otras fuentes, las subjetivas o no autorizas. Puede extraerse mucho conocimiento para investigar a personas, grupos o empresas.
- Consideraciones éticas / legales fuertes. Los resultados descubiertos deben respetar la privacidad de los usuarios y no revelar problemas íntimos y personales. Tener en cuenta las regulaciones relacionadas actuales como el Reglamento General de Protección de Datos. En este sentido, aspectos como la orientación sexual, creencias religiosas, inclinación política o comportamientos comprometedores pueden deducirse de Internet, y este proceso de divulgación puede ser problemático en muchos países. Por otra parte, El alcance de las búsquedas basadas en OSINT debe ser, por definición, limitado a fuentes de datos abiertas.

2.2 Herramientas OSINT

Una de las secciones fundamentales de este trabajo es el estudio exhaustivo de las herramientas OSINT, analizarlas para determinar cuáles son las que cumplen mejor con nuestro propósito. A continuación, se presentan algunas de las principales y más populares herramientas que existen actualmente (*Top 25 OSINT Tools for Penetration Testing*, 2020). Se empieza por las herramientas más globales que abarcan más a la hora de recopilar información, seguido de las más específicas y con un objetivo concreto.

2.2.1 Maltego

Maltego es una de las aplicaciones más potentes y conocidas, desarrollado por la compañía sudafricana llamada Paterva. Es una herramienta de análisis de enlaces de inteligencia de código abierto y lo presenta de manera gráfica para su análisis. Se puede extraer fácilmente datos de fuentes dispersas (registros DNS, registros Whois, motores de búsqueda, redes sociales redes, varias API en línea, metadatos, etc.), fusionar automáticamente información coincidente en un gráfico y mapearla visualmente. Esta herramienta define los siguientes cuatro conceptos principales (Pastor-Galindo et al., 2020).

- Entidad: Es un nodo del gráfico que representa una pieza de información. Algunas entidades predeterminadas son de nombres reales, dirección de correo electrónico,

nombre de usuario, perfil de red social, empresa, organización, sitio *web*, documento, afiliación, dominio, nombre DNS, dirección IP, etc. Además, también podríamos definir entidades personalizadas para nuestra investigación.

- Transformar: Es un código que se aplica a una entidad para descubrir una nueva entidad relacionada. Por ejemplo, se puede aplicar la transformación `` A nombre DNS -MX '' que resuelve un DNS a los registros MX del servidor de correo, podría aplicarse a un dominio nombre de la entidad `` unir.net '' para crear una nueva entidad de registro MX `` unir-net.mail.protection.outlook.com ''. Las transformaciones se podrán ir aplicando recurrentemente propagando el proceso de búsqueda. Además de las transformaciones predeterminadas, también es posible implementar e incluir las propias personalizadas.
- Máquina: Es un conjunto de transformaciones que se definen juntas para ser ejecutado con el fin de automatizar y concatenar muchos procesos de búsqueda.
- Elemento Hub: Es como se les conoce a los diferentes paquetes de transformaciones que los socios agregan al HUB. Al crear tus propias transformaciones integras tus propias fuentes de datos. Esta integración incluye la creación de transformaciones, entidades y máquinas para modelar tus datos.

2.2.2 Recon-ng

Se trata de un *framework* de reconocimiento basado en *web* que funciona por línea de comandos, utiliza un interfaz similar a *Metasploit* donde se configuran opciones y se cargan diferentes módulos. Cada módulo tiene una función diferente, por ejemplo, hay módulos de búsqueda de dominios en *Bing* o *Google*, o módulos de *Shodan*, de perfiles de redes sociales o de cuentas de correo.

Esta herramienta incluye funciones de reconocimiento integradas, interacción con bases de datos, ayuda interactiva y finalización de comandos. Por ejemplo, uno de estos módulos es: *recon/domains-hosts/bing_domain_web*, cargamos este módulo y en el dominio se establece unir.net. Se ejecuta el módulo y se obtiene 71 host encontrados para este dominio.

Toda esta información la va añadiendo a una base local, en la que luego puedes consultar y así realizar la labor de inteligencia a partir de los datos encontrados.

2.2.3 Spiderfoot

Es otra herramienta automática de reconocimiento que recopila información de fuentes públicas. Direcciones IP, redes, dominios, subdominios, direcciones de correos. Los resultados los muestra mediante grafos relacionados entre sí. Permite clasificar la información encontrada por tipo de dato e ir navegando por estos grupos de información, que pueden ser: nombres de dominios, contraseñas comprometidas, correos comprometidos, metadatos, registros DNS, puertos abiertos, etc. Todo dependerá de los datos de entrada que se introduzcan en la herramienta. Por ejemplo, al poner como entrada unir.net obtenemos 215 subdominios.

2.2.4 Foca

FOCA (*Fingerprinting Organizations with Collected Archives*) (*FOCA*, s. f.) es una herramienta diseñada por ElevenPath. Utilizada principalmente para encontrar metadatos e información oculta en los documentos que examina. Estos documentos pueden estar en páginas *web*, y con *FOCA* se pueden descargar y analizar. La cantidad de documentos que es capaz de analizar es muy amplia, principalmente son archivos de Microsoft Office, Open Office, ficheros PDF, ficheros de Adobe InDesign o svg, entre otros.

Para ello emplea motores de búsqueda como son *Google*, *Bing*. Con la suma obtiene gran cantidad de documentos. La aplicación extrae la información oculta de los archivos y los procesa para mostrar al usuario aspectos relevantes. Se descubren, nombres de computadores que crean el documento, la ubicación desde dónde se crearon, los sistemas operativos, nombres y direcciones de correos de usuarios relacionados, fechas de creación, etc. Se puede obtener un mapa de red del objetivo en función a los datos encontrados. Incluye un módulo de descubrimiento de servidores, cuyo objetivo es automatizar el proceso de búsqueda de los mismos usando técnicas enlazadas recursivamente.

El uso de esta herramienta es generalizado en el sector de la seguridad, ya que permite realizar *pentesting* a una empresa. De hecho, puede producir muy buenos resultados debido a que las empresas no suelen limpiar los metadatos de los archivos que se cargan a la red.

2.2.5 Shodan

Motor de búsqueda de dispositivos conectados a internet incluidos los *IoT*. La recolección se hace a través de datos de los servicios HTTP, HTTPS, FTP, SSH, Telnet, SNMP y SIP. Se puede hacer búsquedas por dirección IP, empresas, país o ciudad. Es una herramienta muy usada para la seguridad de red, detectando dispositivos expuestos y sus vulnerabilidades.

2.2.6 The Harvester

Es una herramienta que recopila información pública relacionado con un domino o nombre de empresa. Es capaz de encontrar correos electrónicos, nombres de host, subdominios, direcciones IP y URL relacionados. Permite la exportación a HTML para la representación de resultados. La herramienta se utiliza desde consola de comandos.

2.2.7 Osint Framework

OSINT Framework (*Marco OSINT*, s. f.) es otro tipo de herramienta que se analiza. Realmente se trata un repositorio de herramientas, que de manera gráfica te va guiando los las distintas aplicaciones y técnicas que puedes utilizar en una investigación OSINT. Lo tiene clasificado por diferentes categorías, por donde vas navegando y el árbol de contenido se va ampliando presentándote diversas opciones de aplicaciones. En total son unas treinta y dos categorías principales.

Muchas de las herramientas son *Web* y otras requieren instalación. Esta información lo indica con unos marcadores al lado de la herramienta, para que sea más visual:

- (T) - Indica un enlace a una herramienta que debe instalarse y ejecutarse localmente
- (D) - Google Dorks.
- (R) - Requiere registro
- (M) - Indica una URL que contiene el término de búsqueda y el La URL en sí debe editarse manualmente

Otro repositorio de herramientas muy completo se puede encontrar en (*bellincat*, s. f.) antes mencionada o en (Ph055a, 2018/2020) publicado en Github. También podemos hacer uso de algunas distribuciones Linux orientadas a OSINT en cual incorpora algunas de las herramientas antes mencionadas, como pueden ser: *Osintux*.

Dentro de las numerosas técnicas manuales y servicios disponibles para OSINT destacamos las siguientes:

Búsqueda avanzada en motores de búsqueda

Google y *Bing* permiten búsquedas refinadas haciendo uso de operadores lógicos, por tipos de archivos o títulos de páginas. Por ejemplo, se pueden encontrar gran cantidad de nombres y el DNI del personal educativo de Andalucía con la siguiente búsqueda en Google.

site:juntadeandalucia.es/educacion filetype:pdf intext:dni

También es importante destacar motores de búsqueda específicos para un territorio como *Yandex* que es muy usado en Rusia y el este de Europa. *Baidu* es otro buscador, que es muy usado en Asia.

Búsquedas avanzadas en redes sociales

Facebook es una red social muy extendida y del cual se puede obtener mucha información. Podemos obtener de nuestro objetivo, su lugar de trabajo o puesto, su educación, edad, localización, lugares visitados, gustos. De las fotos publicadas podemos trazar sus actividades.

En *YouTube* a aparte de obtener información de los perfiles, o búsquedas de videos, se pueden encontrar opiniones de los subscriptores. *Twitter* también es interesante para la extracción de opiniones y trazar un perfil. Con *Instagram* se pueden ver localizaciones, personas y actividades realizadas con las imágenes compartidas. *LinkedIn* permite búsquedas por nombre real, empresas, localización, y al ser una red social profesional también se encuentran correos electrónicos y teléfonos.

En este caso también hay redes sociales que son específicas para un determinado país o región. En China las redes sociales más usadas son *Qzone, Weio y Renren*. Este último es muy usado por los universitarios chinos (Lu, 2020). En cambio, en Rusia es muy extendida la red social *VKontakte*, conocida como *VK*.

2.2.8 Comparación de las herramientas

El uso de una u otra herramienta va a depender de los objetivos y necesidades que se tengan a la hora de hacer OSINT. También hay que tener en cuenta que todas las herramientas pueden ser complementarias a la hora de realizar una investigación, y también compaginar con las técnicas manuales. Se ha escogido esta representación de herramientas por considerarlas las principales en su función y utilidad. En la Tabla 2 se puede ver en resumen lo que se destaca de cada una. Si queremos extraer información oculta, usamos *FOCA*. Si nos orientamos a buscar información de red tenemos *Shodan* o *The Harvester*. Por último, para recopilar la máxima información posible está OSINT *Maltego*, *Recon-ng* y *Spiderfoot* que devuelven datos de todo tipo y las relaciones entre ellas. Muy útil para el *pentesting* o la prevención de ataques de *phishing* o ingeniería social.

Tabla 2: Tabla comparativa herramientas OSINT

Herramienta OSINT	**Conjunto de datos de entrada**	**Interfaz**	**Plataforma**	**Resultado a la salida**
Maltego	Información personal, de empresa, dominio	Programa	Linux, Windows, MAC	Múltiple información
FOCA	Dominio, archivos	Programa	Linux, Windows	Metadatos
Recon-ng	Información personal, dominio	Línea de comandos	Linux	Múltiple información
Spiderfoot	Información personal, Información de red, dominio	Interfaz *Web*	Linux, Windows, Online	Múltiple información
The Harverster	Dominio	Línea de comandos	Linux	Información de red. Contactos

Shodan	Cuidad, información de red	Interfaz *Web*	Online	Información de red
Osint Framework	Información personal, de empresa, de red, archivos.	Interfaz *Web*	Online	Múltiple información

Fuente: Basado en (Pastor-Galindo et al., 2020)

2.3 Partes interesadas en la información OSINT

El uso de OSINT tiene mucha relevancia en muchos sectores que pasamos a comentar a continuación (Hassan & Hijazi, 2018):

Gobierno

Especialmente las Fuerzas y Cuerpos de Seguridad del Estado. Muchos son los propósitos como la seguridad nacional, contraterrorismo, ciberataques, comprender la opinión pública nacional y extranjera, información para responsables políticos, o ciberdelitos.

Organismos Internacionales

Por ejemplo, Cruz Roja Internacional emplea OSINT para proteger sus suministros de grupos terroristas. Analizando redes sociales y aplicaciones de mensajería por internet para anticiparse a futuros ataques. También la ONU depende mucho de OSINT a la hora de planificar sus operaciones de paz.

Empresas

La información es poder y las empresas usan las fuentes de OSINT para investigar nuevos mercados, monitorear las actividades de los competidores, realizar el plan de marketing o predecir cualquier evento que pueda afectar a sus estrategias actuales o crecimiento futuro. Otros aspectos que tienen en cuenta las empresas para usar OSINT son las siguientes:

- Protegerse de la fuga de datos, las amenazas cibernéticas se nutren de la exposición de información confidencial y vulnerabilidades en sus redes.

- Junto a un análisis de OSITN interno y externo con otros tipos de información confeccionar una política eficaz de la gestión de riesgos que ayude a proteger sus intereses, financieros, de reputación y datos de los clientes.

***Pentester* y Ciberdelincuetes**

Es una herramienta valiosa para los ataques de ingeniería social. La primera fase de cualquier metodología de prueba de penetración comienza con el reconocimiento, es decir con OSINT.

Organizaciones terroristas

Recopilando información antes de realizar sus ataques, usando mapas o investigando localizaciones, también usando información que ha sido revelada de manera accidental o para desplegar su propaganda por diferentes medios.

Como se puede observar el uso de OSINT es bastante amplio y abarca a todos los ámbitos de la sociedad actual: Seguridad, marketing, reputación online, estudios sociológicos y psicológicos, auditorias y también en el ámbito periodístico. Las campañas de desinformación, *Fake News* o bulos están a la orden del día. Es por ello que el Centro Europeo de Periodismo (EJC) ha publicado guías para la verificación de información utilizando OSINT (*Verification Handbook*, s. f.) o (*Manual de verificación: página de inicio*, s. f.). En estos libros o guías se presentan una serie de herramientas y casos de uso reales para la verificación de información.

En un ambiente competitivo como puede ser la industria farmacéutica (Cerny et al., 2019) hace un análisis del uso de los antidepresivos a nivel europeo. Para ello hace una comparación de la información que pudo encontrar de las agencias de salud y sitios oficiales, con la información que se obtiene *en Google Trends*. Pudiendo constatar que el aumento de uso de estos medicamentos a partir del 2011 también se refleja en el aumento de búsquedas en *Google*.

El Observatorio Europeo de las Drogas y las Toxicomanías (OEDT) elabora un estudio (European Monitoring Centre for Drugs and Drug Addiction, 2019) empleando OSINT para mejorar la monitorización del flujo de la droga que ingresa a Europa. Basándose en la información publicada en medios digitales de cada país sobre incautaciones de más de 100 kilos de cocaína y heroína, pudieron verificar que se corresponde con el tránsito real de la droga que entraba a Europa según sus datos, incluso se detectaron otras vías que no eran consideradas. Pone en relieve la importancia de OSINT y la necesidad e incorporar este tipo de estudios en sus estándares de monitorización.

Las infraestructuras críticas son un blanco para ataques cibernéticos y los daños pueden ser muy graves. En el estudio (Cartagena et al., 2020) se aborda una metodología para el hallazgo de vulnerabilidades mediante OSINT. Un marco en el que los responsables de este tipo de infraestructuras puedan evaluar constantemente la seguridad de sus redes mediante técnicas y herramientas OSINT. En materia de seguridad se puede aplicar de distintas maneras, como proponen (Alves & Ferreira, 2018) de una herramienta que recoge y clasifica *tweets* relacionados con la ciberseguridad publicados por usuarios, empresas de seguridad o ciberdelincuentes y que se puedan aplicar a un tipo de infraestructura TI que se esté monitorizando. Por ejemplo, en el trabajo realizado por (Hernandez Mediná et al., 2018) se analiza una serie de tecnologías OSINT para la labor de ciber inteligencia de la nación y adapta una serie de transformadas al contexto colombiano.

En definitiva como menciona (Fantinelli & Sivilli, 2015)

"la implementación de la metodología Open Source Intelligence (OSINT) dentro de la gestión organizacional puede fortalecer la reputación de la marca y las actividades de inteligencia competitiva. Representa un instrumento de amplio alcance para recopilar datos e información para que las organizaciones mejoren la toma de decisiones, realicen análisis preventivos de riesgos, mejoren los procesos de adquisición de información de diligencia debida, supervisen la efectividad de la comunicación organizacional y la reputación en línea."

2.4 Estado actual de la digitalización en el ámbito educativo

Como en otros ámbitos el sector educativo está inmerso en la transformación digital, en el informe (*La transformación digital del sector educacion*, s. f.) de la Fundación Orange, destaca que la diferencia más importante con otros sectores es la coexistencia de entornos públicos y privados, el cual supone tendencias y grupos de interés muy diferentes. La transformación afecta plenamente al sector educativo y formativo, por lo que empresas e instituciones educativas están adaptando tanto sus contenidos, objetivos y los medios a un mundo digital.

Las previsiones de *Global Center for Digital Business Transformation*, elabora un ranking de industrias según vemos en la siguiente tabla, y posiciona en este informe (*The Digital Vortex in 2019*, s. f.) a la industria de la Educación en el octavo lugar, es decir en un punto intermedio.

Tabla 3: Ranking de industrias según su potencial de disrupción digital

Industria	Ranking
Medios y Entretenimiento	1
Productos y Servicios Tecnológicos	2
Telecomunicaciones	3
Retail	4
Servicios Financieros	5
Turismo	6
Transporte y Logística	7
Educación	**8**
Servicios Profesionales	9
Productos de embalaje de consumo	10
Salud y Farmacéuticas	11
Manufactura	12
Energía y utilidades	13
Construcción	14

Fuente: (*The Digital Vortex in 2019*, s. f.)

Según *Global Market Insight* (*E-Learning Market Trends 2020-2026 | Global Research Report*, s. f.) el mercado del *e-Learning* superó los 200 mil millones de dólares en el 2019, y como se ve en la Figura 6 se espera que crezca un 8%, gracias a la llegada de varias tecnologías, como la computación en la nube e inteligencia artificial junto al crecimiento de acceso a internet en el mundo que impulsará el mercado. En concreto el mercado europeo dominó con más del 35% de ingresos y mantendrá su crecimiento. Además, el despliegue de la tecnología 5G facilitará la capacitación y el aprendizaje sin interrupciones.

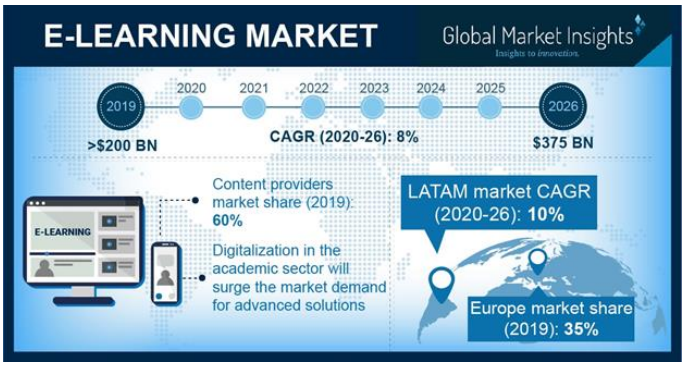


Figura 6: Tendencia del mercado e-Learning

Fuente: (*E-Learning Market Trends 2020-2026 | Global Research Report*, s. f.)

La oferta global de *MOOC* (*Massive Online Open Courses*) también están al alza. En 2019, se identificó un total de 11.450 cursos para ser brindados por varios proveedores de MOOC en todo el mundo, un 22% más que el año pasado (*MOOC Market Forecast, Trend Analysis & Competition Tracking - Global Market Insights 2019 to 2029*, s. f.).

Como dato que refleja el impulso de este tipo de formación, según (*Mercado MOOC por Plataformas y Servicios - 2023* | *Mercados y Mercados*, s. f.):

- En octubre de 2018, *edX* presentó 9 nuevos programas de master de las principales instituciones mundiales en temas altamente demandados, como ciencia de datos, ciberseguridad, informática, análisis y gestión de la cadena de suministro.
- En marzo de 2018, *Coursera* se asoció con 5 universidades para ampliar sus ofertas de maestría y licenciatura para impulsar una mayor base de estudiantes.

En España según el (*Dossier de Indicadores de la Sociedad Digital por género (marzo 2020)* | *Ontsi - Red.es*, s. f.) en el año 2019 un 31% de la población total ha realizado alguna actividad de educación y formación en línea. Si nos situamos en los centros educativos no universitarios según la Tabla 4 del (*Dossier de Indicadores relacionados con la administración electrónica y las TIC en la educación en España (febrero 2019)* | *Ontsi - Red.es*, s. f.), el número medio de alumnos por ordenador destinado a tareas de enseñanza y aprendizaje es de 3.

Tabla 4: Número de ordenadores en los centros educativos

	Total	Público	Privado
Número medio de alumnos por ordenador destinados a tareas de enseñanza y aprendizaje	3	2,8	3,6
Número medio de alumnos por ordenador con conexión a Internet destinado a la docencia	3,8	3,5	4,5
Número medio de alumnos por ordenador destinado a la docencia con alumnos	3,6	3,3	4,2
Número medio de profesores por ordenador	1,9	1,8	2,1

Fuente: (*Dossier de Indicadores relacionados con la administración electrónica y las TIC en la educación en España (febrero 2019)* | *Ontsi - Red.es*, s. f.)

La totalidad de los centros educativos poseen conexión a internet, y el 86% tiene una página *Web* publicada en Internet. Respecto a conexión WIFI se reflejan los datos en la siguiente tabla:

Tabla 5: Porcentaje de uso WIFI en los centros educativos

Centros educativos con conexión wifi	90
Con servicios de Internet	86
Alumnos con dispositivos del centro	72
Alumnos con dispositivos propio	17

Fuente: (*Dossier de Indicadores relacionados con la administración electrónica y las TIC en la educación en España (febrero 2019)* | *Ontsi - Red.es*, s. f.)

Respecto a las plataformas educativas online y servicios en la nube, el 37% del total de los centros públicos poseen un entorno virtual de aprendizaje y un 46% poseen servicios en la nube, podemos ver en la Tabla 6 los datos diferenciados por tipo de centro.

Tabla 6: Porcentaje de plataformas educativas online y servicios en la nube

	Centros privados	Centros públicos enseñanza primaria	Centros públicos enseñanza secundaria y F.P.	Total, centros públicos
Centros educativos con servicios de entorno virtual de aprendizaje	49	25	68	37
Centros educativos con servicios en la nube	64	41		46

Fuente: (*Dossier de Indicadores relacionados con la administración electrónica y las TIC en la educación en España (febrero 2019)* | *Ontsi - Red.es*, s. f.)

Por otro lado, en la Tabla 7 se reflejan estos datos por tipo de proveedor, es decir en el sector público se tiende a utilizar las plataformas de aprendizaje que dota la administración, y en el sector privado son ajenos al centro y la administración. En cambio, para los servicios en la nube hay una preferencia tanto de los centros públicos y privados de los servicios externos al centro y la administración.

Tabla 7:Servicios de entorno virtual y en la nube por tipo de proveedor

		Servicios externos ajenos al centro y a la administración educativa	La administración educativa	El propio centro
Centros educativos con servicios de entorno virtual de aprendizaje (EVA) por tipo de proveedor	Publico	19	61	26
	Privado	51	9	47
Centros educativos con servicios en la nube por tipo de proveedor	Público	39	33	28
	Privado	59	2	39

Fuente: (*Dossier de Indicadores relacionados con la administración electrónica y las TIC en la educación en España (febrero 2019)* | *Ontsi - Red.es*, s. f.)

La estadísticas del Ministerio de Educación (*Estadísticas de la Educación*, s. f.) que reflejamos en la Tabla 8, nos ayuda a poner en contexto los datos expuestos anteriormente. El número de alumnos no universitarios en España es de más de 8,2 millones y de casi 1,6 millones de universitarios. Si a esto le sumamos el personal docente tenemos un total de 10.6 millones de personas entre alumnos y docentes en España, un 22,5% del total de la población.

Tabla 8: Número de alumnos en España

Indicador	Periodo	valor
Alumnado Enseñanzas Régimen General no universitarias	2018-2019	8.217.651
Alumnado Enseñanzas Universitarias (Grado, 1º y 2º ciclo, Máster y Doctorado)	2018-2019	1.595.039

Profesorado Enseñanzas Régimen General no universitarias	2018-2019	712.181
Personal Docente e Investigador E. Universitaria	2017-2018	122.910

Fuente: (*Estadísticas de la Educación*, s. f.)

El Plan Digital 2020 elaborado por la CEOE (*Plan Digital 2020: la digitalización de la sociedad española*, s. f.) establece como uno de los pilares fundamentales priorizar la transformación del sistema educativo para adecuarlo a la nueva Sociedad Digital. Esto refuerza la idea de que la digitalización de los Centros Educativos está en constante crecimiento.

Estos datos ponen de manifiesto que la cantidad de información es enorme y que estas instituciones son un objetivo importante para el ciberdelito. Según un estudio de Ciberseguridad en el sector universitario llevado a cabo por *Deloitte* (*Las ciberamenazas ponen en alerta a las universidades*, s. f.) el 80% de las universidades que participaron declararon haber sufrido algún incidente en los 12 meses.

En un estudio realizado por *Hiscox* (Criado, 2020), la compañía de seguros especializado en ciberseguridad, destaca las seis principales amenazas y objetivos para este año 2020, uno de estos objetivos en que han puesto el foco los ciberdelincuentes es el ataque a instituciones, como son las administraciones públicas y las instituciones educativas.

Ya en el 2018 (*El sector educativo, un blanco perfecto para ciberataques*, s. f.) se menciona que el sector educativo se clasifica como uno de los más susceptibles al riesgo cibernético. Esto es debido al tipo de información que albergan y por el tipo de entorno tienden a ser las redes más fáciles de penetrar. Destaca que los datos personales que albergan son muy valiosos, son datos de direcciones, números de cuenta, información médica, empleados, proveedores, etc. Las instituciones educativas como hemos visto fomentan el aprendizaje en línea y en persona, y a sus redes ingresan todo tipo de usuarios con diferentes tipos de dispositivos que no pueden ser monitoreados todo el tiempo. Los estudiantes por ejemplo utilizan smartphones para acceder a las redes sociales, y muchas instituciones carecen de políticas de seguridad para seguir el ritmo a los factores de riesgo inherentes en la evolución tecnológico.

Según podemos ver en la Figura 7 las filtraciones de datos por industria, el sector educativo ocupa el cuarto lugar (Portillo, 2019).

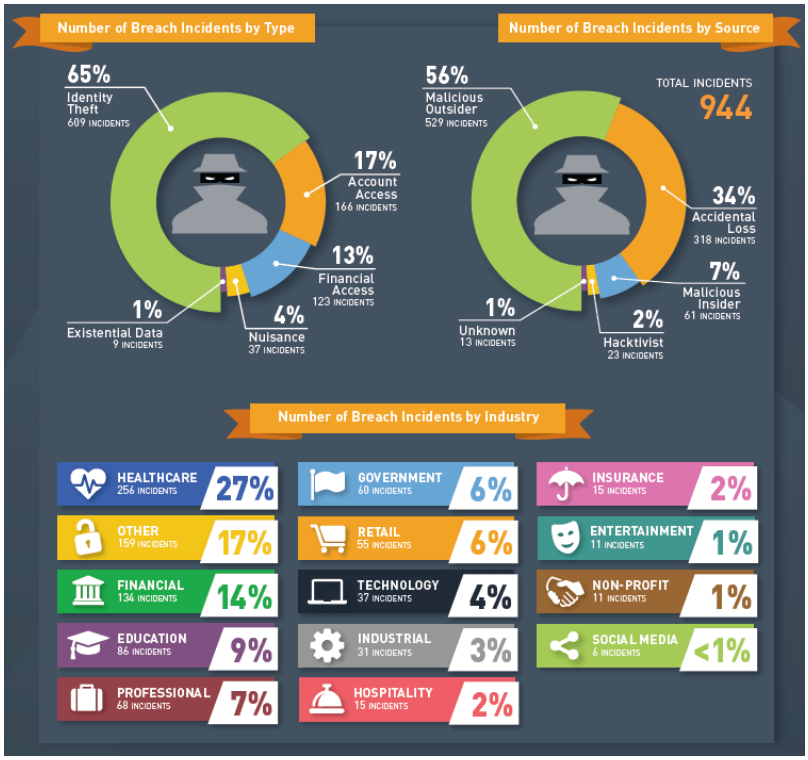


Figura 7: Filtraciones de datos por industria

Fuente: (Portillo, 2019)

Todo este cambio hacia lo digital se ha visto drásticamente impulsado por la llegada de la pandemia provocada por el SARS-CoV 2. Según la UNESCO (UNESCO, 2020) 1.200 millones de estudiantes están fuera de las aulas debido al cierre de los colegios y universidades por la pandemia del SARS-CoV-2. Los sistemas electrónicos como por ejemplo *Google Classroom*, entre otros, están ayudando a esta conexión a distancia de los alumnos y profesores. Los esfuerzos nacionales a gran escala para aprovechar la tecnología para el apoyo de la educación a distancia y aprendizaje durante la pandemia están evolucionando rápidamente (*E-Learning Market Trends 2020-2026 | Global Research Report*, s. f.), además se espera que la industria *e-Learning* tenga un crecimiento mayor de lo estimado, sobre todo en países más duramente golpeados por la pandemia como Italia, España y Francia.

En este nuevo contexto muchas de las instituciones no estaban preparadas y se han visto desbordadas. La improvisación, falta de formación e inexperiencia ha puesto en riego los datos de alumnos y docentes (Martín-Arroyo, 2020) amenazando la privacidad de los menores. Incluso al extremo de crear grupos de *Whatsapp* o *Telegram* para tener un contacto directo con los alumnos.

2.5 Conclusión estado del arte

A lo largo de este capítulo se ha descrito qué es OSINT y cómo funciona. Se identifican cuáles son sus puntos fuertes y también los problemas a los que nos enfrentamos al emplearlo. Todo este proceso requiere de una metodología, unos pasos que hay que seguir para finalmente producir inteligencia, que es el objetivo.

Se destacan las fases de recopilación que se dividen en identificar las fuentes y la adquisición de información propiamente dicha, para luego realizar un procesamiento y análisis con el objetivo final de producir inteligencia o extraer conocimiento. Estas fases se apoyan en herramientas y técnicas que ayudan y automatizan estos procesos. Se hace una selección de las herramientas más relevantes y se destacan las técnicas más empleadas, para luego estudiarlas y compararlas.

También se han visto los distintos sectores donde se emplea OSINT, desde el gobierno e instituciones hasta empresas de todo tipo, además de las diferentes motivaciones de emplear esta técnica. Para finalmente llegar al apartado anterior y conocer el estado actual de la digitalización en el ámbito educativo a nivel global y en particular a nivel nacional en España. Concluyendo que el sector educativo está en gran expansión hacia la digitalización y cada vez la información generada en digital va en aumento. Teniendo en cuenta el número de personas que hay en la comunidad educativa (Tabla 8). Es de destacar como dato representativo que actualmente el 68% de los centros públicos de educación secundaria en España tienen una plataforma virtual de aprendizaje (Tabla 6).

En Andalucía, esta plataforma virtual de aprendizaje de los centros públicos se llama *Moodle Centros*. Además, existe una aplicación para la comunicación del centro educativo con las familias llamado *PASEN*, con versión *web* y para dispositivos móviles. Y por último la aplicación *Séneca*, para que el profesorado y el centro educativo pueda llevar a cabo todo el proceso de gestión administrativa que conlleva la labor docente. Los centros tienen páginas *Web*, *blogs* y muchos tienen cuentas en redes sociales. Hay mucha información sensible que hay que proteger, que quizás con el mal uso de la tecnología y malas prácticas no se está atendiendo.

Estudiadas las herramientas y los estudios realizados en otros ámbitos, hacen comprender la potencia de estas técnicas, lo poco explotado en algunos sectores y el amplio recorrido que tiene por delante. El uso de OSINT aplicado al ámbito educativo es interesante porque son instituciones que tienen ciertas peculiaridades, se manejan datos personales sensibles y en la mayoría de los casos de menores de edad. Mi aportación se centra en analizar las

herramientas aplicándolas a un objetivo dentro del sector educativo. De esta manera determinar cuáles de estas herramientas encajan mejor para este ámbito. A la conclusión de este análisis se da respuesta a con cuáles de estas herramientas se pueden realizar mejores auditorías, prevención de fuga de datos, detectar la exposición de información o vulnerabilidades en las redes. Este estudio debe servir como parte de los análisis preventivos de riesgos y guía de mejores prácticas.

3. Objetivos concretos y metodología

3.1 Objetivo General

El objetivo principal de este trabajo fin de master es dotar a las instituciones educativas de unas herramientas para poder emplear OSINT y valorar sus datos expuestos con el fin de protegerse.

3.2 Objetivo específico

Con el presente trabajo se pretende alcanzar estos objetivos

- Analizar OSINT, conocerlo en profundidad y los diferentes estudios al respecto. Así como sus metodologías y aplicaciones, ventajas e inconvenientes.
- Explorar las tendencias de OSINT. Las aplicaciones más relevantes y que se puedan alinear con nuestro objetivo.
- Conocer el estado actual de las instituciones educativas en el ámbito digital y comprender sus necesidades.
- Clasificar las diferentes Herramientas OSINT y medir cuales o que parte de ellas encajan mejor para realizar OSINT a una institución educativa.
- Realizar un análisis con estas herramientas a una institución educativa.
- Proponer las herramientas o parte de ellas que funcionen mejor para realizar OSINT en una institución educativa.

3.3 Metodología de trabajo

La metodología del trabajo a realizar se podría englobar en cuatro fases principales:

Estudio del estado del arte

Investigación documental para la recolección de fuente de referencia bibliográfica, reportes, libros, sitos *webs* de interés, así como investigación científica por parte de los expertos, las

líneas futuras y los problemas en el campo de OSINT. Este estudio es pilar fundamental para establecer que temas son relevantes, son generalidades o antecedentes o líneas de trabajo futuras que pueden tener cierto paralelismo con las instituciones educativas.

Estudio instituciones educativas

Ver cuál es el estado actual y las tendencias en las instituciones educativas, en su digitalización, nos permitirá saber, cuáles serán sus objetivos, sus retos, riesgos y oportunidades. Además, nos ayudará a ver los puntos en común con otros sectores. En función de las necesidades específicas para este ámbito y otros sectores que comparten escenarios parecidos, decidir qué objetivos son los más importantes a cumplir.

Propuesta de piloto experimental

Una vez alcanzado esto, podemos ver que herramientas, previamente estudiadas, son las que cumplen mejor estas condiciones para que sea aplicable a este tipo instituciones. Consistirá en aplicar OSINT en un objetivo concreto del sector educativo, teniendo en cuenta sus particularidades y los objetivos que se quiere alcanzar.

Estudio de resultados

Por último, obtenidos los resultados de aplicar el proceso OSINT. Se contrastarán y valorarán los resultados obtenidos con el fin de determinar cuáles de estas herramientas o partes de ellas, presentan mejor rendimiento.

4. Descripción del experimento

En este capítulo se elegirá un objetivo del ámbito educativo al que aplicar OSINT, se configurarán las herramientas antes estudiadas y se realizarán las pruebas con cada herramienta sobre nuestro objetivo.

El objetivo es un centro público de educación secundaria y de ciclos formativos, de una localidad de Andalucía. Se elige este tipo de institución por varios motivos, entre ellos por temas puramente que lo distinguen de otros tipos de organización, según (Saballs, 2005).

- Son organizaciones que se plantean muchos objetivos.
- Las tareas de los educadores son múltiples y la división de trabajo es poco clara.
- Desde dentro tiene diferentes puntos de vistas de cómo deben funcionar.
- Escasez de recursos.
- Débil articulación de la organización, no siempre todo queda regulado.
- El poder errático de los directores. La elección sobre personas sin ningún perfil previo comporta un modelo de dirección desprofesionalizado.
- Los recursos asignados les llegan, fundamentalmente, a través de decisiones de naturaleza política.

Desde otras perspectivas, enumeramos las razones de esta elección, frente a otro tipo de instituciones educativas.

- En Andalucía hay en total 1.608.790 alumnos no universitarios matriculados en el curso 2018 – 2019. De los cuales 1.184.243 son en centros públicos.
- De estas cifras, los centros de educación secundaria y ciclos formativos tienen 514.226 alumnos matriculados.
- Este tipo de instituciones, manejan en su totalidad datos de medio millón de menores. Estos datos son de especial protección y además pueden ser muy sensibles. Datos académicos, de salud, económicos, asuntos sociales, direcciones, teléfonos, etc.
- Estas instituciones son muy heterogéneas y la manera de trabajar cambia mucho de una a otra.
- En su mayoría, los alumnos son adolescentes con acceso a internet y redes sociales, lo que los convierten en un objetivo muy vulnerable.
- A diferencia de las universidades, son organizaciones mucho más pequeñas, lo que supone menos recursos.

- Con el estado de alarma debido a la pandemia, hemos pasado a una educación online, en la que los datos de estos menores están más expuestos aún, debido a la improvisación, falta de conocimientos y recursos. Lo que ha llevado a utilizar herramientas poco recomendables. Como pueden ser las redes sociales.

Por estas razones, nuestro objetivo es un Instituto de Educación Secundaria que se encuentra en un municipio de unos 80.000 habitantes, en el que hay otros tres institutos de educación secundaria más. Es un centro educativo de tamaño medio por el número de alumnos y representa un objetivo muy estándar de una institución de este tipo en Andalucía y en todo el territorio nacional.

4.1 Entorno de trabajo

El entorno de trabajo desde el cual se realizan las tareas de OSINT debe seguir unos criterios mínimos en materia de seguridad, básicamente son:

- Sistemas actualizados: sistemas operativos, antivirus, y software utilizado.
- Utilizar contraseñas robustas.
- Utilizar unidades cifradas.
- No usar el ordenador personal.
- Usar navegación anónima por internet.
- Seguridad en email y redes sociales.

Para la ejecución de las pruebas se hace uso de los siguientes equipos:

- VMware Workstation: Para no usar el ordenador personal empleando la virtualización. Se instala la versión 15.5.
- Máquina virtual Kali Linux: Versión Kali 5.6.0 x86_64 GNU/Linux. Kali viene con muchas de las aplicaciones instaladas, por ejemplo: *Recon-ng* y *The Harvester*.
- Máquina Virtual Windows 10: En esta máquina virtual se instala *Maltego* y *FOCA*.

Las herramientas que se van a probar son las que se han estudiado en el estado de arte. Con cada herramienta se harán todas las pruebas que nos permita obtener resultados en sus versiones gratuitas. Y para ello partiremos de la misma información. Es decir, partiendo del domino o dirección *web* del objetivo, se ejecutarán las herramientas una a una.

4.2 Pruebas con Maltego

Como se ha comentado en el capítulo 2 todo el entorno es gráfico, lo que hace bastante intuitivo y muy rápido empezar a ejecutar las pruebas. La versión que se utiliza para estas pruebas es *Maltego Community Edition 4.2.9*, requiere hacerse una cuenta gratuita y ya puedes tener acceso a gran parte de sus opciones.

Se crea un nuevo proyecto y en la ventana del nuevo trabajo ya puedes empezar a arrastrar entidades. Las pruebas las empezamos con la entidad dominio, como se observa en la Figura 8. Ponemos el nombre del domino, que es el de nuestro objetivo. A partir de esta información vamos a ir realizando transformadas para ir detectando, recopilando toda la información que podamos y nos permite usar la herramienta.

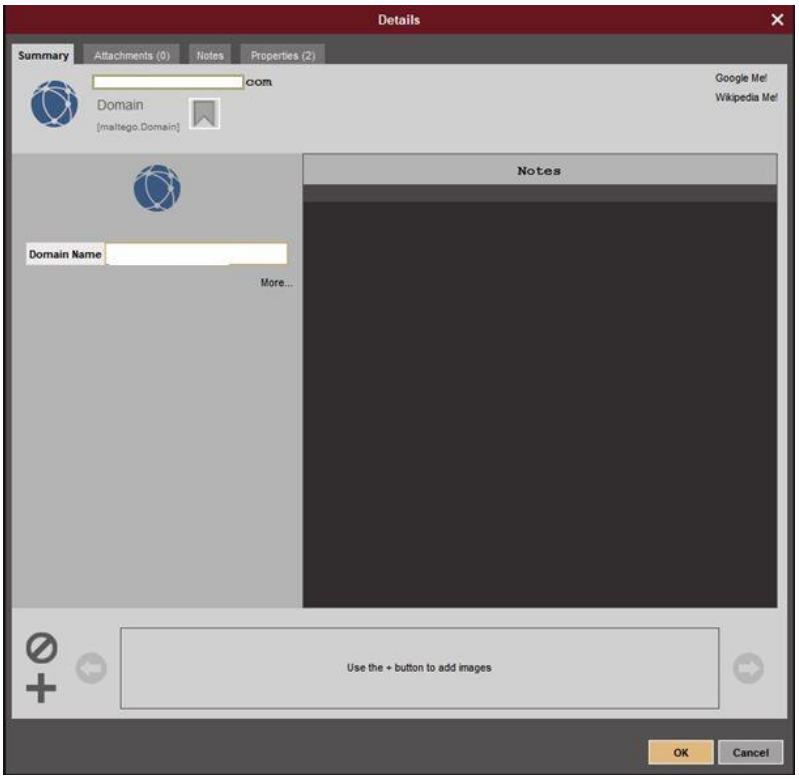


Figura 8: Nombre del dominio

A partir de esta entidad origen, se van encadenando transformadas según se vaya viendo la información relevante que se va encontrando, en la Tabla 9 se detallan las transformadas utilizadas para este objetivo. Para realizar la transformada, hay que pulsar con el ratón derecho sobre una entidad y te aparecen todas las transformadas que se pueden realizar.

Tabla 9: Transformadas realizadas en maltego

Transformada	Entidad Origen	Propósito
To entities from WHOIS	Dominio	Ver whois del dominio
To DNS Name - NS	Dominio	Encontrar nombres de los servidores del dominio
To DNS Name	Dominio	Encontrar subdominios
To DNS Name - MX	Dominio	Encontrar registro MX del dominio
To email @domain	Dominio	Encontrar direcciones de correo con el dominio
To website	Dominio	Encontrar la página *web*
Mirror: Email address	Página *Web*	Encontrar direcciones de correo electrónico
To IP Address	Página *Web*	Encontrar IP
To Server Technologies	Página *Web*	Encontrar la tecnología que usa para la página *Web*
To Domain [DNS]	Dirección de correo electrónico	Encontrar el nombre del dominio
To DNS Name - MX	Dominio (otro encontrado)	Encontrar sus registros MX
Mirror: External Links	Página *Web /docs.google*	Encontrar *links* a sitios externos
Verify email address exists	Correo electrónico	Comprobar que el correo existe
Mirror email address	Facebook/ twiiter/ docs.google	Encontrar direcciones de correo electrónico

4.3 Pruebas con Recon-ng

Se abre la consola y se ejecuta el comando *recon-ng*, se puede ver en la Figura 9 el interfaz de la aplicación, que tiene la versión *recong-ng V5.0.1* en el que se observa que dispone de 83 módulos. A partir de este momento se puede empezar a ejecutar los comandos para ir cargando los módulos que se vayan a utilizar. Antes de nada, se crea un espacio de trabajo donde se van recopilando toda la información obtenida, lo llamamos TFM y se ejecuta el siguiente comando

```
[recon-ng][default] > workspaces create TFM
```

Ya se tiene el proyecto creado, se puede listar los *workspaces* creados con este comando:

```
[recon-ng][default] > workspaces list
```

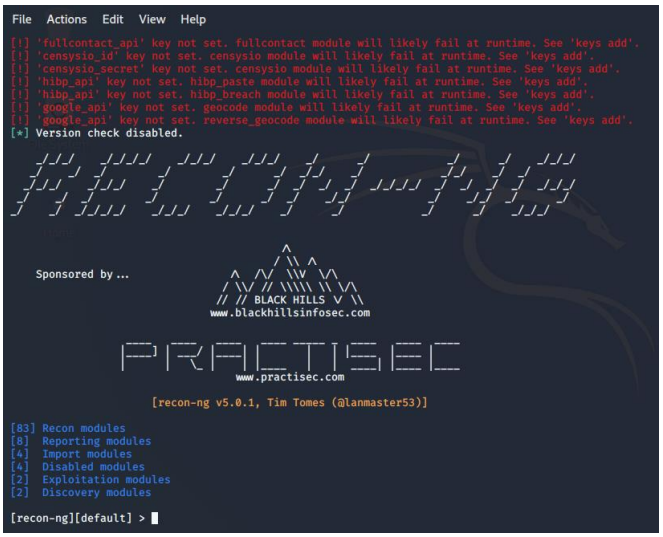
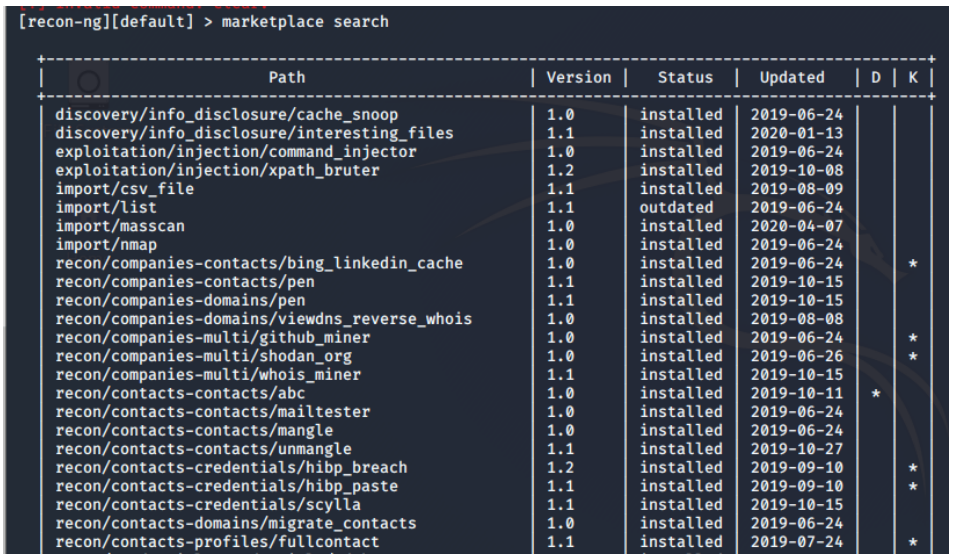


Figura 9: Ejecutar aplicación recon-ng

Al ejecutar el siguiente comando se puede ver la lista de todos los módulos disponibles en esta aplicación, la columna D significa que ese módulo tiene dependencias y la columna K significa que se necesita una clave para su ejecución, se puede ver en la Figura 10.

```
[recon-ng][default] > marketplace search
```



Path	Version	Status	Updated	D	K
discovery/info_disclosure/cache_snoop	1.0	installed	2019-06-24		
discovery/info_disclosure/interesting_files	1.1	installed	2020-01-13		
exploitation/injection/command_injector	1.0	installed	2019-06-24		
exploitation/injection/xpath_bruter	1.2	installed	2019-10-08		
import/csv_file	1.1	installed	2019-08-09		
import/list	1.1	outdated	2019-06-24		
import/masscan	1.0	installed	2020-04-07		
import/nmap	1.0	installed	2019-06-24		
recon/companies-contacts/bing_linkedin_cache	1.0	installed	2019-06-24		*
recon/companies-contacts/pen	1.1	installed	2019-10-15		
recon/companies-domains/pen	1.1	installed	2019-10-15		
recon/companies-domains/viewdns_reverse_whois	1.0	installed	2019-08-08		
recon/companies-multi/github_miner	1.0	installed	2019-06-24		*
recon/companies-multi/shodan_org	1.0	installed	2019-06-26		*
recon/companies-multi/whois_miner	1.1	installed	2019-10-15		
recon/contacts-contacts/abc	1.0	installed	2019-10-11	*	
recon/contacts-contacts/mailtester	1.0	installed	2019-06-24		
recon/contacts-contacts/mangle	1.0	installed	2019-06-24		
recon/contacts-contacts/unmangle	1.1	installed	2019-10-27		
recon/contacts-credentials/hibp_breach	1.2	installed	2019-09-10		*
recon/contacts-credentials/hibp_paste	1.1	installed	2019-09-10		*
recon/contacts-credentials/scylla	1.1	installed	2019-10-15		
recon/contacts-domains/migrate_contacts	1.0	installed	2019-06-24		
recon/contacts-profiles/fullcontact	1.1	installed	2019-07-24		*

Figura 10: Algunos módulos de recon-ng

Para cargar un módulo solo hay que ejecutar el comando seguido del módulo, por ejemplo, para cargar el módulo de whois:

```
[recon-ng][TFM] > modules load recon/domains-contacts/whois_pocs
```

Para ver los campos y modificar las opciones dentro del módulo cargado, se ejecuta el siguiente comando:

```
[recon-ng][TFM][whois_pocs] > info
```

Como se ve en la Figura 11, con este comando se observa una pequeña descripción del módulo y las opciones que tiene, en este caso se tiene que poner un dominio para que haga su tarea. Para modificar una opción, en este caso el dominio, se usa el siguiente comando:

```
[recon-ng][TFM][whois_pocs] > options set SOURCE DominioEjemplo.com
```

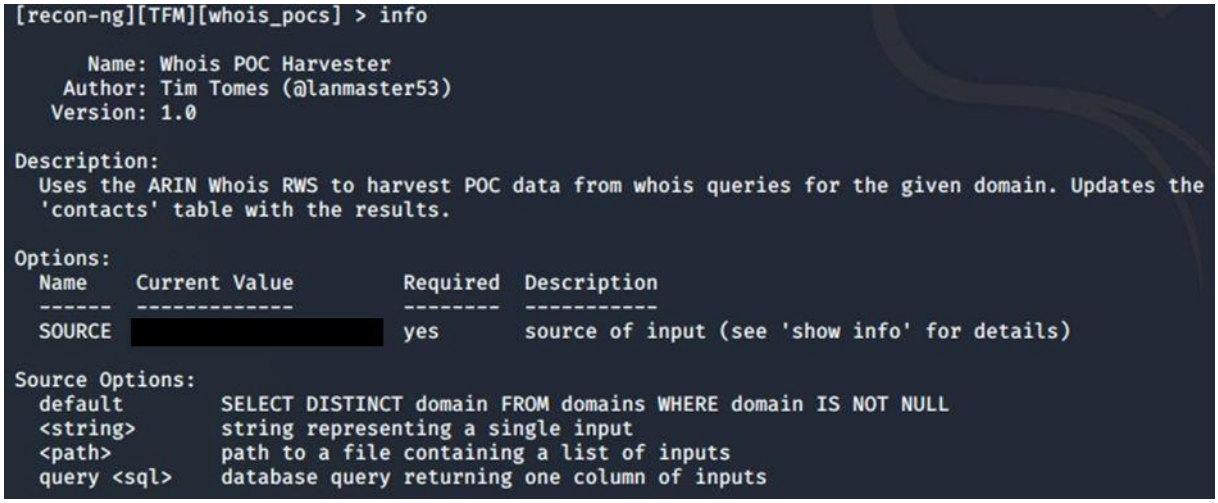


Figura 11: información módulo recon-ng

En la Figura 12 se detallan los módulos usados con esta herramienta, para realizar las pruebas sobre nuestro objetivo.

```
[recon-ng][TFM] > dashboard

+-------------------------------------------------------------+
|                       Activity Summary                      |
+-------------------------------------------------------------+
|                         Module                          | Runs |
+-------------------------------------------------------------+
| discovery/info_disclosure/interesting_files             | 1    |
| recon/contacts-contacts/mailtester                      | 1    |
| recon/domains-contacts/hunter_io                        | 1    |
| recon/domains-contacts/whois_pocs                       | 1    |
| recon/domains-credentials/pwnedlist/domain_ispwned      | 1    |
| recon/domains-hosts/bing_domain_web                     | 1    |
| recon/domains-hosts/findsubdomains                      | 1    |
| recon/domains-hosts/google_site_web                     | 1    |
| recon/domains-hosts/hackertarget                        | 1    |
| recon/domains-hosts/mx_spf_ip                           | 1    |
| recon/domains-hosts/netcraft                            | 1    |
| recon/domains-vulnerabilities/xssed                     | 1    |
| recon/profiles-contacts/bing_linkedin_contacts          | 2    |
| recon/profiles-profiles/profiler                        | 1    |
+-------------------------------------------------------------+
```

Figura 12: Módulos ejecutados para la prueba

4.4 Pruebas con Spiderfoot

Esta herramienta en su versión web *SpiderFoot HX*, solo es necesario registrarse y de manera gratuita se puede hacer uso de gran parte de sus opciones.

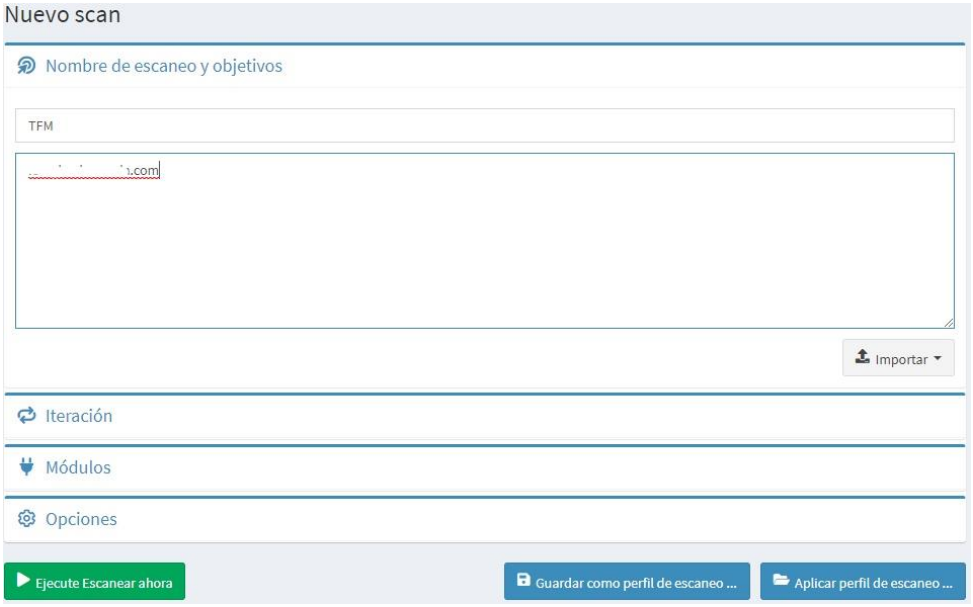


Figura 13: Configuración escaneo spiderfootHX

Cuando se selecciona un nuevo escaneo, como se observa en la Figura 13 se puede añadir el objetivo de diversas formas, es decir, puedes añadir el dominio, nombres de *host*, IP, número de teléfonos, emails, nombres de usuarios o nombres reales de personas. Te permite añadir varias de ellas a la vez separadas por un espacio. En nuestro caso se añade el dominio.

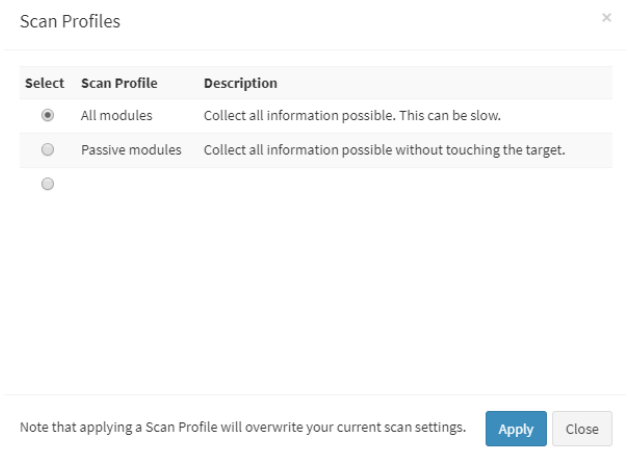


Figura 14: Usar todos los módulos

En la opción de módulos que se ve en la Figura 14, se marca la opción que utilice todos los módulos, para así obtener la mayor información posible. Después de realizar esto ya se está en disposición de ejecutar el escaneo. La aplicación también tiene la opción de monitorizar, con esta opción se puede programar tareas de escaneo a un objetivo y que nos envíe las notificaciones por email.

4.5 Pruebas con Foca

Se ejecuta la aplicación, se crea el proyecto y se añade la dirección *web* del dominio. En la Figura 15 se refleja la creación del proyecto donde se añade la dirección del dominio del objetivo.

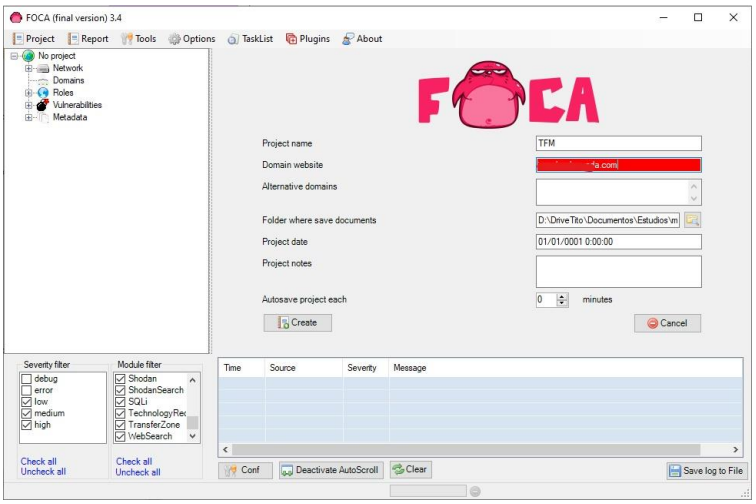


Figura 15: Creación proyecto FOCA

Hecho esto y alguna configuración previa, Figura 16. Se lanza la prueba de esta aplicación sobre el objetivo.

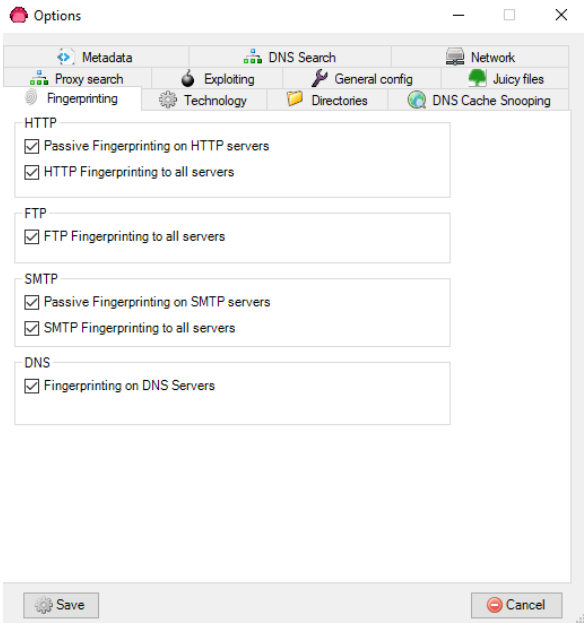


Figura 16: Configuración FOCA

4.6 Pruebas con Shodan

La búsqueda es sencilla, en este caso por el tipo de objetivo se tiene solo acceso a una IP, como se observa en la Figura 17 se hace esta búsqueda por IP en la *web* de *Shodan.*

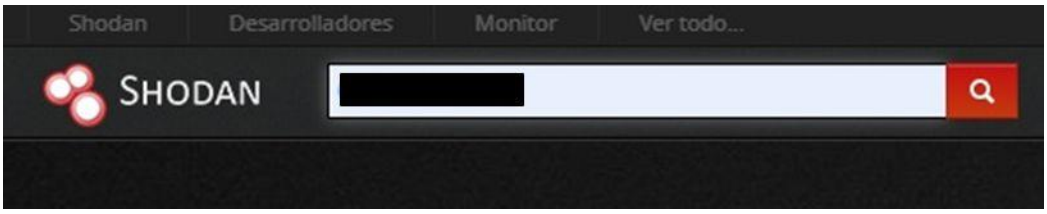


Figura 17: Busqueda en Shodan

4.7 Pruebas con The Harvester

Con esta herramienta se hace la búsqueda con un solo comando sobre nuestro objetivo. Sobre la consola se ejecuta el siguiente comando para hacer una búsqueda. Con estos parámetros se intenta realizar búsquedas del dominio objetivo en todos los orígenes de datos.

```
kali@kali: ~$ theHarvester -d DominioEjemplo.com -b all
```

4.8 Pruebas con Repositorios de herramientas

Con el uso de un repositorio, se pierde la automatización del proceso. Se utiliza OSINT Framework y se hace uso de herramientas que puedan dar información del domino y de redes sociales. En la Figura 18 se observa la variedad y cantidad de herramientas que propone usar *OSINT Framework*.

Se proponen una serie de herramientas *web* que pueden dar información a partir del dominio, como:

- whois.domaintools.com
- pentest-tools.com
- spyse.com
- buitwith.com

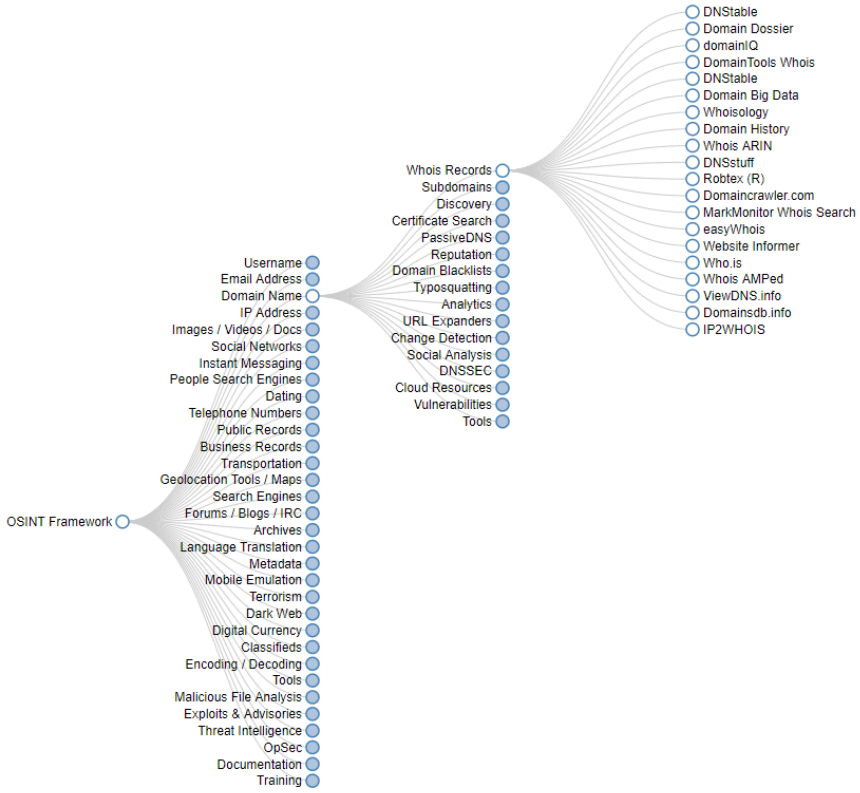


Figura 18: Variedad de herramientas en OSINT Framework

5. Presentación de los resultados

Una vez realizadas las pruebas que se describen en el capítulo anterior, a lo largo de esta sección se presentarán los resultados obtenidos con cada una de las herramientas.

5.1 Resultados con Maltego

Según la Figura 19 por ejemplo, para la transformada *WHOIS*, se dibuja el nodo principal que es el dominio y la información que ha encontrado y cuelga de él.

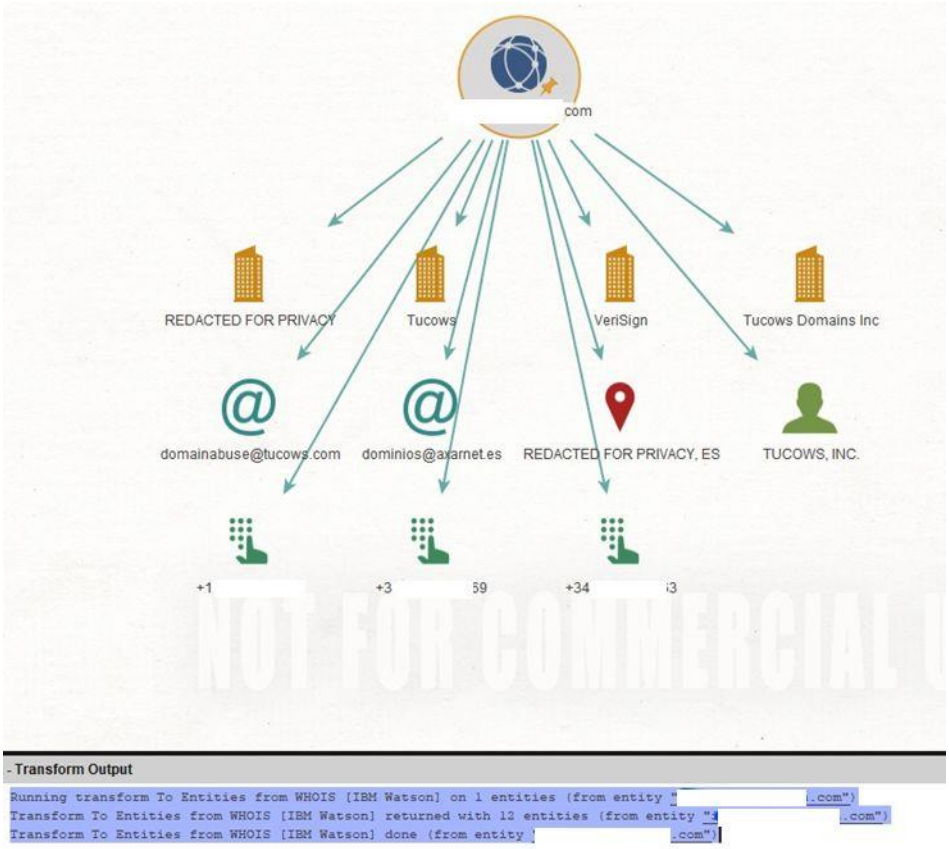


Figura 19: Información obtenida con la transformada WHOIS

Pero para valorar mejor los resultados finales de esta prueba, empleamos la opción de exportar los resultados a un informe PDF de 39 páginas. El cual nos recopila todo lo que ha sido capaz de encontrar.

De manera general se puede resumir esta información clasificando las entidades encontradas por tipos. En la Figura 20 se puede ver la tecnología por el cual está construido el dominio, los subdominios encontrados, que en este caso son tres y también los dominios relacionados, en el que encuentra un segundo dominio.

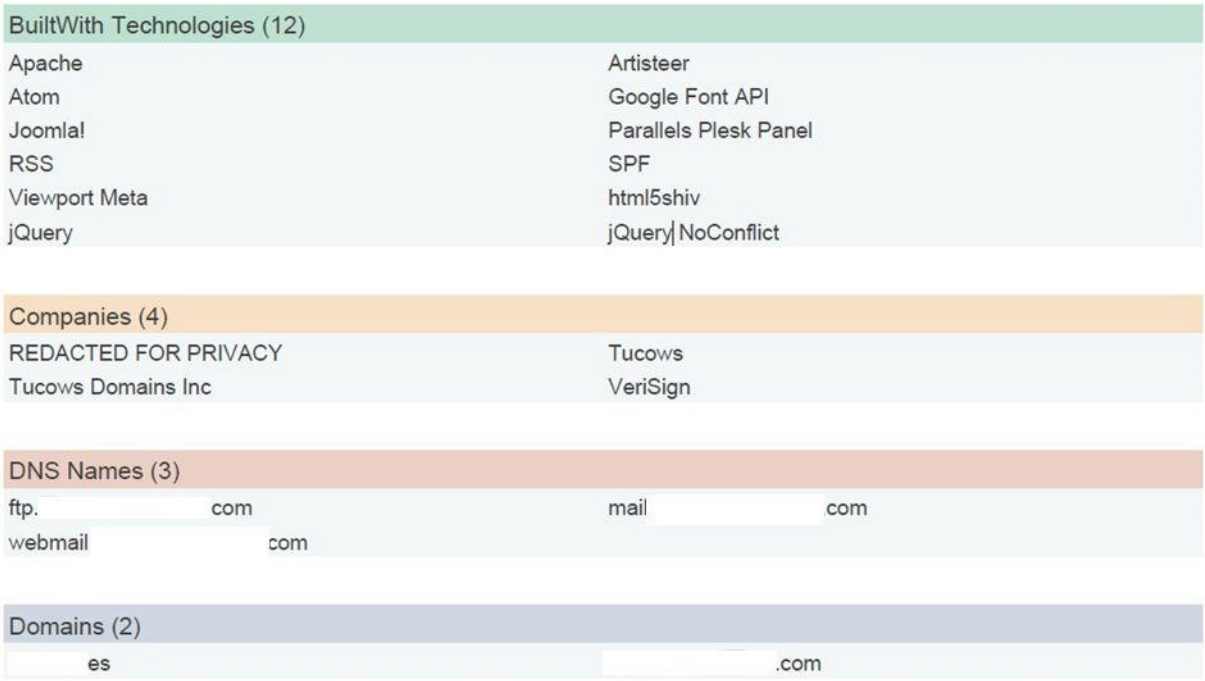


Figura 20: Información del dominio

En la Figura 21 se observan los emails encontrados, algunos encontrados en la *web* y otros desde el Facebook del objetivo, evidentemente aquí hay muchos datos que habría que descartar o analizar con más detenimiento.

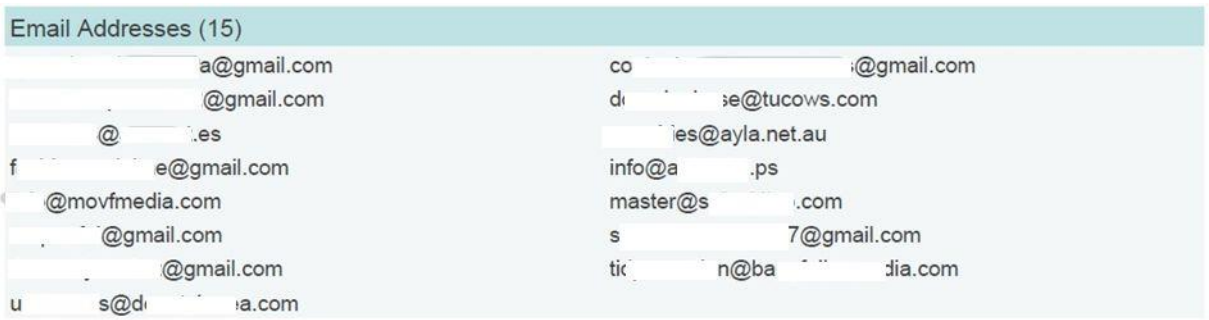


Figura 21: Emails encontrados

En la Figura 22 se hace referencia a la IP y los registros MX encontrados.

Figura 22: IP y registros MX

Otra de la información encontrada es los registros del nombre de servidor, números de teléfonos y enlaces a páginas *web*. En la Figura 23 destacamos páginas de documentos de Google, a las redes sociales, algún blog y páginas a la Junta de Andalucía donde residen las plataformas para alumnos y profesores.

Figura 23: Nombres de servidor y enlaces

El informe, para cada entidad ofrece mayor detalle en los apartados siguientes del mismo. Por ejemplo, para el enlace a documentos de Google especifica todas las URL encontradas, como

se observa en la Figura 24. Lo miso hace con el enlace a la Junta de Andalucía que especifica todas las URL.

Figura 24: Enlaces de documentos Google

5.2 Resultados con Recon-ng

A continuación, se muestra la información recopilada con cada módulo. En general se ha encontrado poca información con esta herramienta. En muchos de los módulos es necesario una clave de las API y con otros módulos obtenemos cero resultados. Por ejemplo, con *findsubdomains*, *whois-pocs*, *netcraft*, *xssed*, *Bing-domain-web* o *Google-site-web* no se obtienen resultados. Otros como *mailtester, hunter-io, domain-ispwned, Bing-linkedin-contacts* requieren clave para la API. Y la idea es ejecutar estas aplicaciones sin otro recurso adicional.

En la Figura 25 se observa que el módulo *hackertarget* ha recopilado o encontrado tres hosts para este dominio.

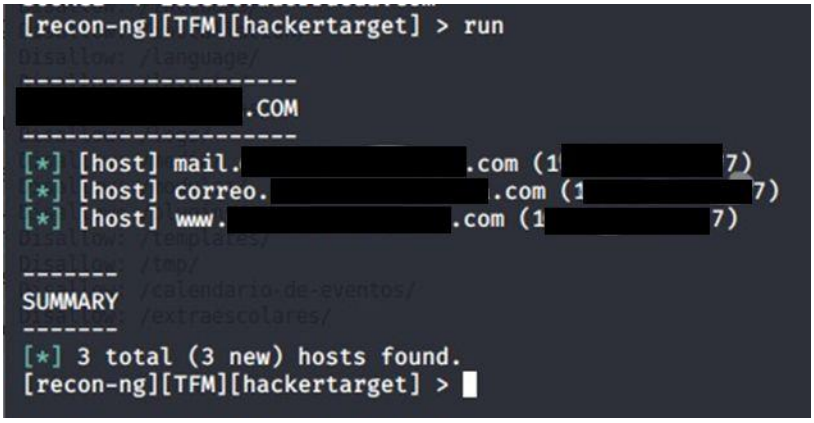


Figura 25: recon-ng módulo hackertarget

El módulo *interesting_files* solo ha encontrado el archivo robots.exe. Por otro lado, en la Figura 26 podemos ver que el módulo *mx-spf-ip* encuentra los archivos MX y el módulo *profiler* (Figura 27) encuentra perfiles para este dominio.

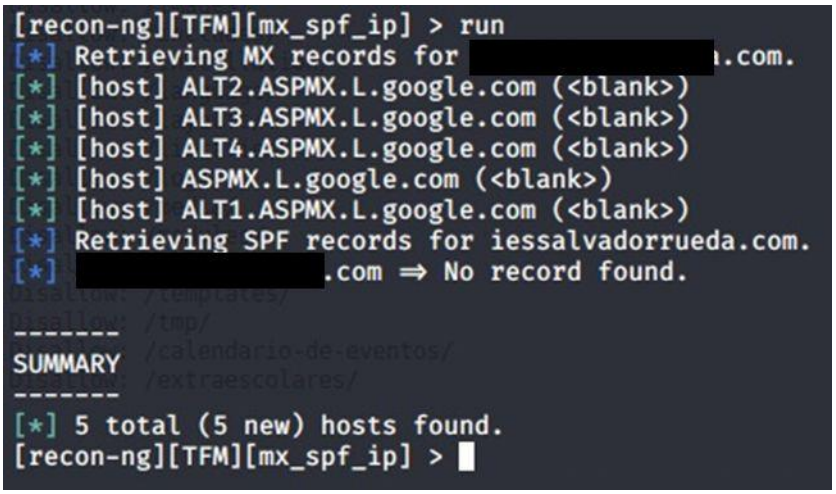


Figura 26: recon-ng módulo mx spf ip

Figura 27: recon-ng módulo profiler

5.3 Resultados con Spiderfoot

En el gráfico de la Figura 28, muestra como visión general toda la información que ha encontrado. Dice que en total son 17998 elementos de datos, de los cuales 12012 son datos únicos. Por otro lado, muestra una información de las correlaciones clasificado por nivel de riesgo y tipo, también se ve a golpe de vista un gráfico de los módulos, sus categorías y la cantidad de datos encontrados por cada uno de ellos. Otro de los gráficos muestra el origen de estos datos y la cantidad de ellos que ha encontrado. Por último, en la parte inferior muestra todos los tipos de datos que ha encontrado y la cantidad de información de cada uno de ellos.

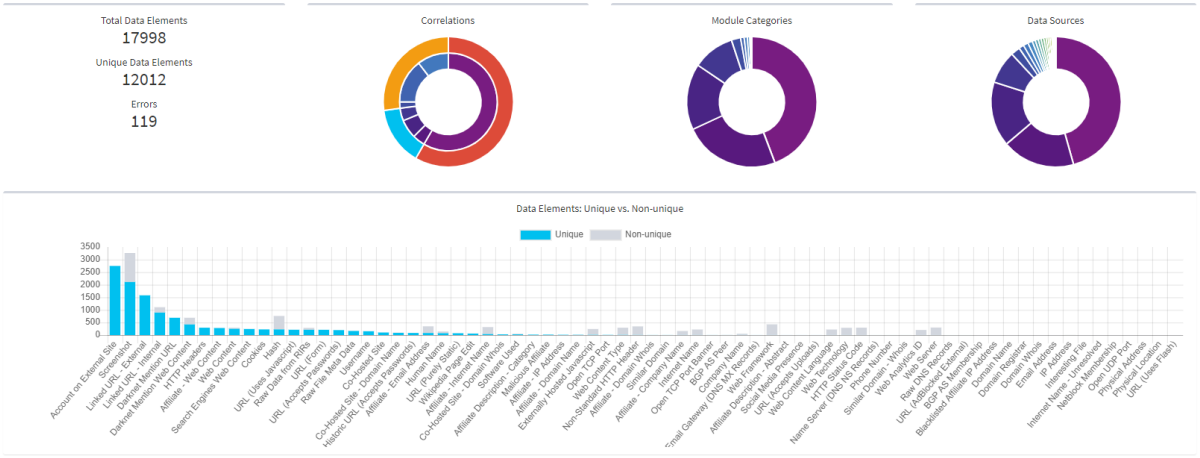


Figura 28:spiderfootHX visión general

Si se clasifica la información por tipo de dato: En la Figura 29 se ve parte de esta información, en la primera columna te muestra el tipo de datos, en otra columna el número de datos por cada tipo y en la segunda columna, en rojo, te marca si en lo encontrado existe un riesgo, se pueden destacar varios tipos. Por ejemplo, el tipo *Co-Hosted site*, el cual permite ver a 103 host que comparten IP.

Affiliate – Domain Name nos da mucha información de dominios relacionados con el objetivo, sobre todo blogs. En el tipo *Affiliate – Email Address*, es capaz de presentar muchas direcciones de email, aunque muchos de ellos no resultan relevantes. Esta herramienta hace búsquedas en la *Darknet,* hay muchos resultados que en un principio no parecen relevantes.

Data Type	Risky	Unique	Total	% of Data
Account on External Site	0	2749	2749	15%
Affiliate - Company Name	0	6	162	1%
Affiliate - Domain Name	0	15	24	0%
Affiliate - Domain Whois	0	7	7	0%
Affiliate - Email Address	0	83	349	2%
Affiliate - IP Address	0	17	17	0%
Affiliate - Internet Name	0	39	319	2%
Affiliate - Web Content	0	280	292	2%
Affiliate Description - Abstract	0	4	4	0%
Affiliate Description - Category	0	22	24	0%
BGP AS Membership	0	1	2	0%
BGP AS Peer	0	5	11	0%
Blacklisted Affiliate IP Address	1	1	1	0%
Co-Hosted Site	0	103	103	1%
Co-Hosted Site - Domain Name	0	88	103	1%
Co-Hosted Site - Domain Whois	0	29	33	0%
Company Name	0	5	54	0%
Cookies	0	226	226	1%
Darknet Mention URL	689	689	689	4%
Darknet Mention Web Content	423	423	689	4%

Figura 29: SpiderfootHX clasificación por tipo de datos

La Figura 30, refleja la Información referente a lo que se ha encontrado en las tecnologías utilizadas por el servidor.

Web Content Type	0	12	294	2%
Web Framework	0	5	428	2%
Web Server	0	3	300	2%
Web Technology	0	4	290	2%
Wikipedia Page Edit	0	61	61	0%

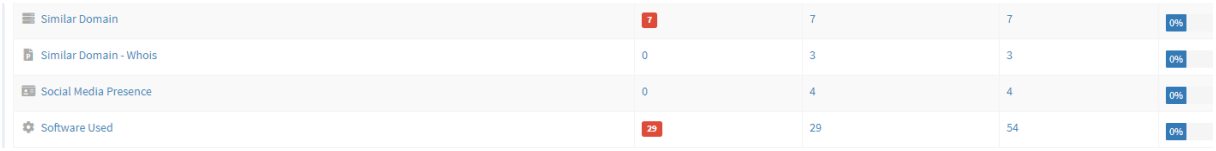
Figura 30: SpiderfootHX tecnologías utilizadas

En la Figura 31 vemos información relativa al dominio, relativa a los registros MX, Host externos con *Javascript* y cabeceras HTTP. Destacamos *Email Adddress* al encontrar una cuenta de correo electrónico perteneciente al dominio.

Domain Name	0	1	2
Domain Registrar	0	1	1
Domain Whois	0	1	1
Email Address	0	1	2
Email Gateway (DNS MX Records)	0	5	5
Externally Hosted Javascript	0	14	241
HTTP Headers	0	294	294
HTTP Status Code	0	3	296

Figura 31: SpiderfootHX información del dominio

Otro tipo de datos que nos da es el nombre de dominios similares al nuestro, que además lo califica con riesgo elevado. También la presencia en redes sociales, enumerando las redes sociales a las que ha encontrado para este dominio. Otra información relevante es el software usado. Para detectarlo, la aplicación ha podido acceder a esta información a través de los metadatos de los archivos o enlaces, pudiendo detectar software de edición de imagen o de tecnologías. Todo lo mencionado lo vemos en la Figura 32.



Similar Domain	7	7	7	0%
Similar Domain - Whois	0	3	3	0%
Social Media Presence	0	4	4	0%
Software Used	29	29	54	0%

Figura 32: SpiderfootHX dominios similares y redes sociales

La Figura 33 muestra información relevante: La aplicación encuentra nombres de personas. No encuentra todos los nombres y muchos de ellos no son a tener en cuenta porque no son nombres reales. Los enlaces a páginas externas es otro dato relevante, aquí hay más de 1500, es necesario hacer una limpieza de estos datos, pero da muy buena información, enlaces a otras *web*s, blogs, páginas de la Junta de Andalucía, redes sociales, documentos compartidos por Google.

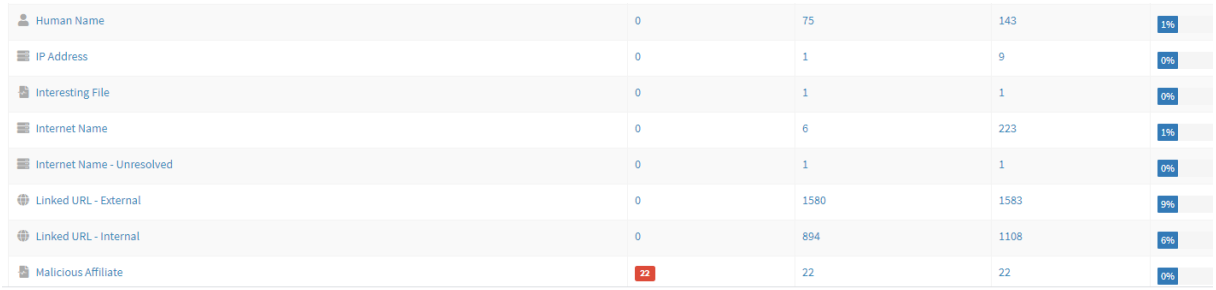


Figura 33: SpiderfootHX nombres de personas y enlaces externos

El escaneo de puertos detecta 13 puertos TPC abiertos e incluso un puerto UDP como se ve en la Figura 34.

Figura 34: SpiderfootHX puertos abiertos

Hay otras formas de visualización de los datos, y también te puedes descargar CSV para hacer tus propias búsquedas, descartar datos y cruzar datos. Existe otra opción de trabajar con esta herramienta, que es la investigación, a raíz de una entrada como el dominio, vas desarrollando tu búsqueda de manera gráfica y paso a paso, parecido a *Maltego*.

5.4 Resultados con Foca

Con *FOCA* se encuentra información del dominio, direcciones IP, la tecnología que utiliza, puertos y servicios abiertos, como podemos ver en la siguiente figura. También encuentra un segundo dominio relacionado con nuestro objetivo. Respecto a documentos, archivos y su análisis de metadatos no encuentra nada. Los documentos en realidad son enlaces *html* y no hay prácticamente ningún archivo subido a la *web* por esta razón no da ningún resultado.

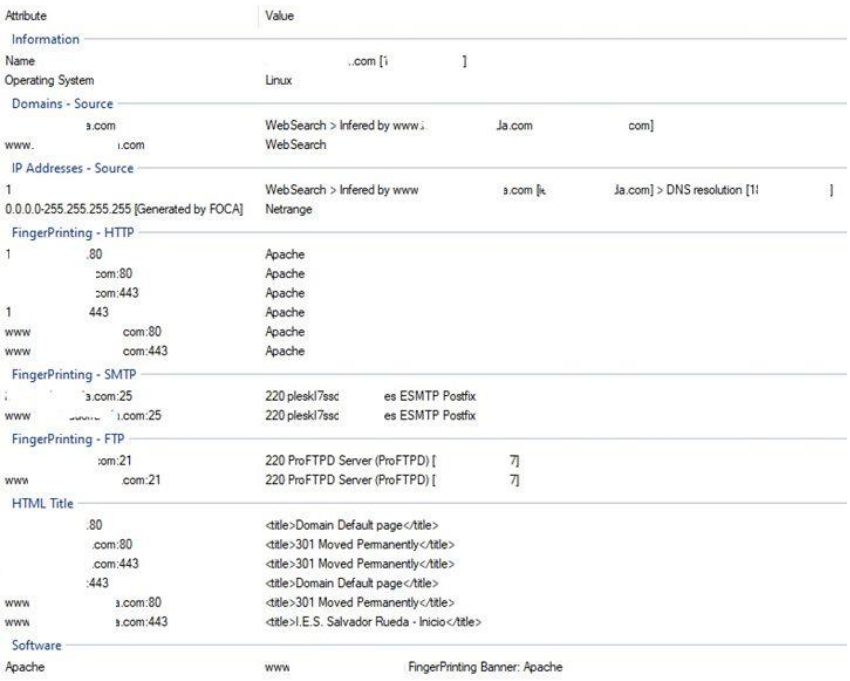


Figura 35: FOCA resultados dominio

5.5 Resultados Con Shodan

En la Figura 36, se pueden ver que con *Shodan* se obtiene información de los puertos abiertos, la tecnología usada, la plataforma sobre la que se desarrolla esta tecnología, nombres de *host* donde están alojados estos servicios.

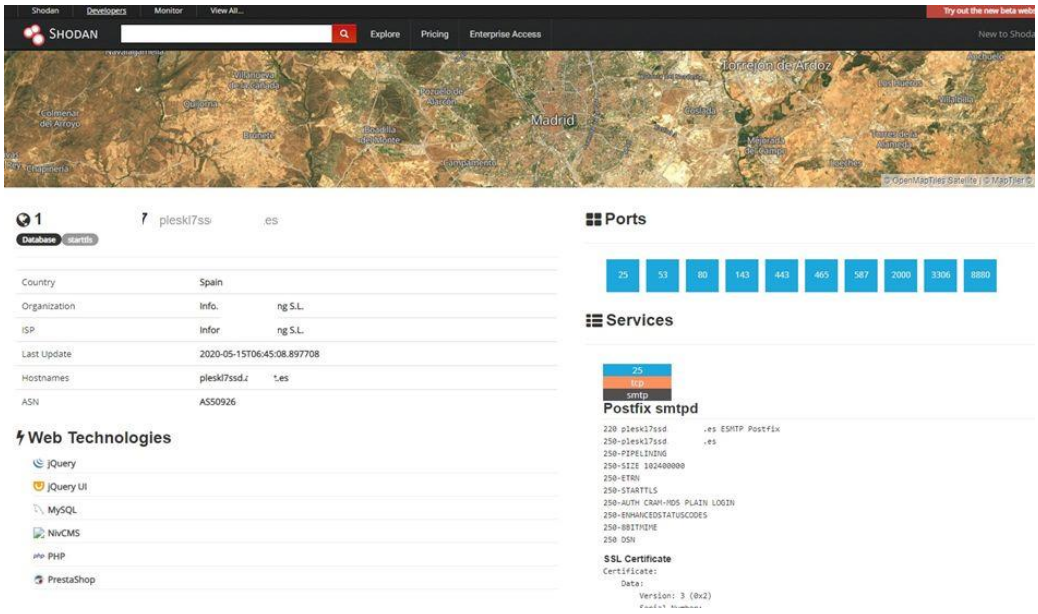


Figura 36: Shodan resultados encontrado IP dada

5.6 Resultados con The Harvester

En la Figura 37 se muestra parte de los resultados obtenidos. Realmente para este objetivo se encuentra muy poca información con esta aplicación. Se localizan algunos resultados en *Linkedin*.

Figura 37: TheHarvester resultados

5.7 Resultados con Repositorios de herramientas

A continuación, se detallan los resultados que se obtienen con estas herramientas *Web* de escaneo de dominios. En su conjunto se obtienen los siguientes resultados.

whois.domaintools.com

En esta *Web* se obtienen datos de IP, localización, proveedor de servicio, y tipo de servidor. Además, información del dominio *Whois*. Se pueden ver en las Figura 38 y Figura 39.

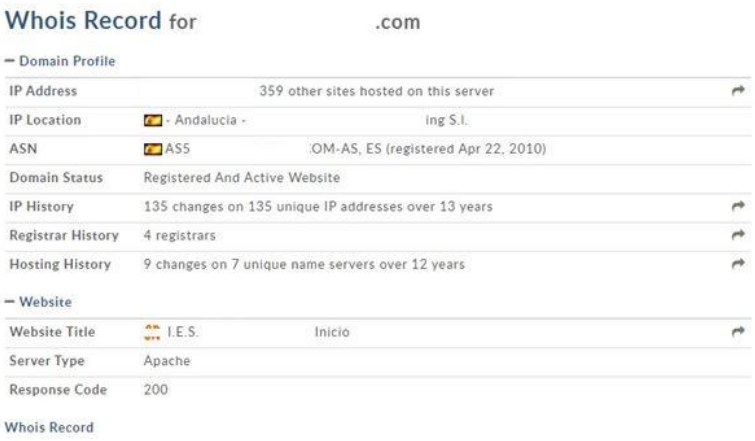


Figura 38: Whois.domaintools.com resultado 1

```
Domain Name:                A.COM
Registry Domain ID: 809477970_DOMAIN_COM-VRSN
Registrar WHOIS Server: whois.tucows.com
Registrar URL: http://tucowsdomains.com
Updated Date: 2020-01-14T09:43:42
Creation Date: 2007-02-11T19:14:26
Registrar Registration Expiration Date: 2021-02-11T19:14:26
Registrar: TUCOWS, INC.
Registrar IANA ID: 69
Reseller: A       SL
Domain Status: ok https://icann.org/epp#ok
Registry Registrant ID:
Registrant Name: REDACTED FOR PRIVACY
Registrant Organization: REDACTED FOR PRIVACY
Registrant Street: REDACTED FOR PRIVACY
```

Figura 39: Whois.domaintools.com resultado 2

pentest-tools.com

Con esta aplicación lo que se encuentra son los subdominios, como podemos observar en la Figura 40.

Subdomain	IP address	OS	Server	Technology	Web Platform	Page Title
.com			Apache	PHP 7.0.33, PleskLin		I.E.S. - Inicio
www.i om	1		Apache	PHP 7.0.33, PleskLin		I.E.S. - Inicio
ftp .com	1 7		Apache			Domain Default page
webma .com	7		Apache	PHP 5.4.16, PleskLin		Roundcube Webmail :: Welcome to Roundcube Webmail
mai com	1		Apache			Domain Default page

Figura 40: Pentest-tools.com resultados

spyse.com

El escaneo que realiza esta aplicación, permite encontrar los siguientes datos: IP, registro PTR, el proveedor de servicio, registros MX, y registros NS. Podemos verlo en las Figura 41 y Figura 42.

A registros - 1

IP	Registro PTR	Puntuación Spyse	Geo IP	ASN	Organización AS	ISP
	pleskl7s es	60 60	España	50926	In g SL	Infort

Figura 41: Spyse.com resultado

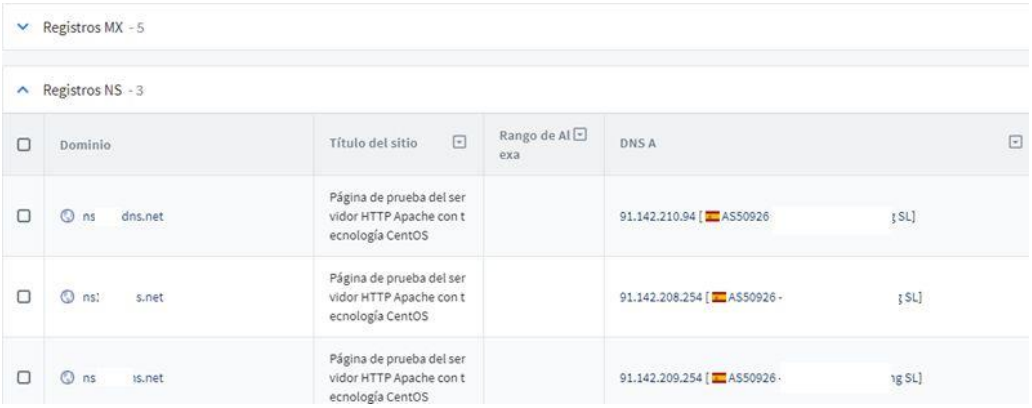


Figura 42: Spyse.com resultado registros

builtwith.com

Con esta aplicación encontramos información interesante de respecto a otros dominios que comparten IP, además del histórico de IP del domino investigado. Se puede ver parte de esta información en la Figura 43.

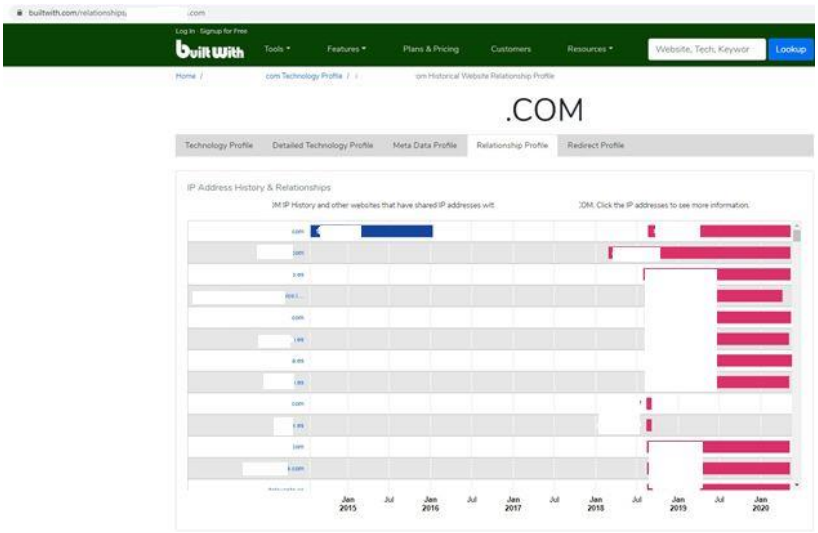


Figura 43: Buitwith.com resultados

6. Discusión de los resultados

A lo largo de este capítulo se realizará un análisis de los resultados mostrados en el capítulo anterior. Estableciendo comparativas en los puntos más relevantes, donde se verán las ventajas, inconvenientes y puntos fuertes de cada una de las herramientas que se han empleado. Para finalizar, se expondrá brevemente una serie de recomendaciones a los centros educativos, para evitar la sobreexposición de información detectada en este tipo de instituciones, que ponen en riesgo los datos de los centros educativos.

6.1 Análisis de resultados

Antes de empezar el análisis por el tipo de información encontrada por cada herramienta, en la Figura 44 se plasma el tiempo, en minutos, que ha empleado cada herramienta en la ejecución de las pruebas. Se aprecian claramente estas diferencias en la figura siendo *Spiderfoot* la que más tiempo requiere y *shodan* la que menos. *Maltego* y *Recon-ng* necesitan de nuestra intervención para dar el siguiente paso. Sin embargo, el resto de herramientas las pruebas se han realizado automáticas de principio a fin, es decir una vez configurado se lanza la ejecución y se espera a que finalicen.

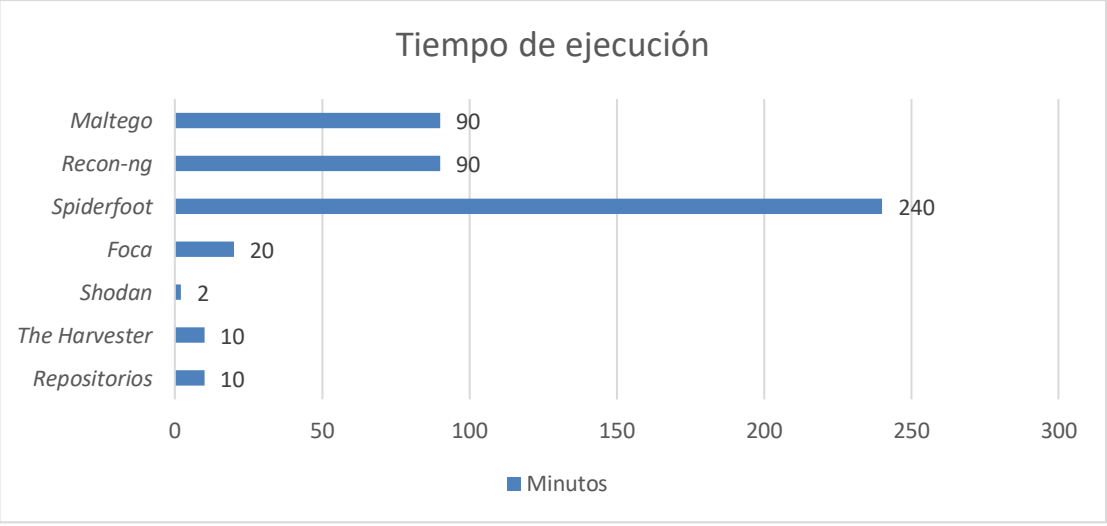


Figura 44: Tiempo de ejecución de las herramientas

6.1.1 Datos del domino

En la Tabla 10, se puede ver una comparación de los resultados obtenidos con las herramientas en este aspecto. Con *Maltego*, *Spidefoot* y los repositorios se obtiene toda la información respecto a la tecnología utilizada por el servidor, los registros MX y NS, además de encontrar subdominios y la empresa de *hosting*.

Tabla 10: Comparación datos del dominio

	Tecnología del servidor	Registros MX	Registro NS	Subdominios	Empresa Hosting
Maltego	Si	Si	Si	Si	Si
Recon-ng	No	Si	No	Si	No
Spiderfoot	Si	Si	Si	Si	Si
Foca	Si	No	No	No	Si
Shodan	Si	No	No	No	Si
The Harvester	No	No	No	No	No
Respositorios	Si	Si	Si	Si	Si

Se destaca que con *Spiderfoot* se logra encontrar un subdominio más como se puede observar en la Figura 45. Poder recopilar esta información con estas herramientas era esperable, aunque se esperaba poder hacerla también con *Recon-ng* con el que solo ha sido posible encontrar información parcial. *FOCA* y *Shodan* da buena información resumida respecto a la tecnología utilizada. También es cierto, que por la información que da de puertos abiertos, podemos deducir que presta algún servicio en un subdominio.

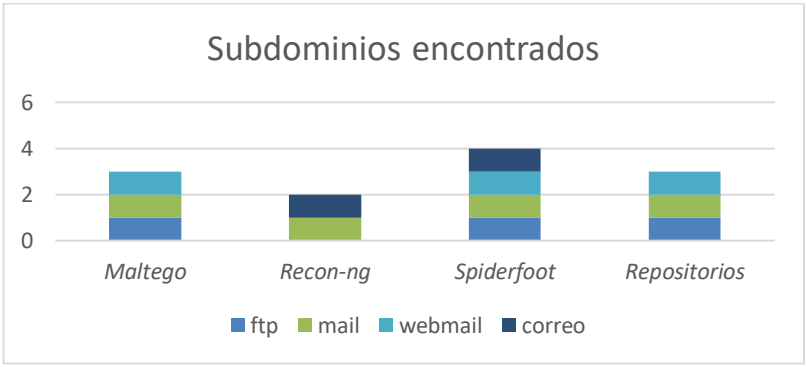


Figura 45: Subdominios encontrados por cada herramienta

6.1.2 Emails y enlaces

Con *Maltego* y *Spiderfoot* se encuentra información en todos los campos analizados como se ve en la Tabla 11.

Tabla 11: Comparación emails, enlaces y teléfono

	Emails	Enlaces a redes sociales	Enlaces a otras *webs*	Número de teléfonos
Maltego	Si	Si	Si	Si
Recon-ng	No	Si	No	No
Spiderfoot	Si	Si	Si	Si
The Harvester	No	No	No	No

En menor medida, se encuentran datos con *Recon-ng* y con *the Harvester* ninguna información. Respecto a los emails, *Maltego* localiza la dirección principal publicada en la *web* y algunas direcciones más de correo que en su mayoría no son relevantes según vemos en la Figura 46. En cambio, con *Spiderfoot* a parte del correo principal publicado en la *web*, encuentra una dirección de correo del propio dominio que *Maltego* no localiza, además de otras muchas más cuentas de correo, del orden de cuatro veces más, lógicamente muchas de ellas no aportan información útil, pero otras si son de interés.

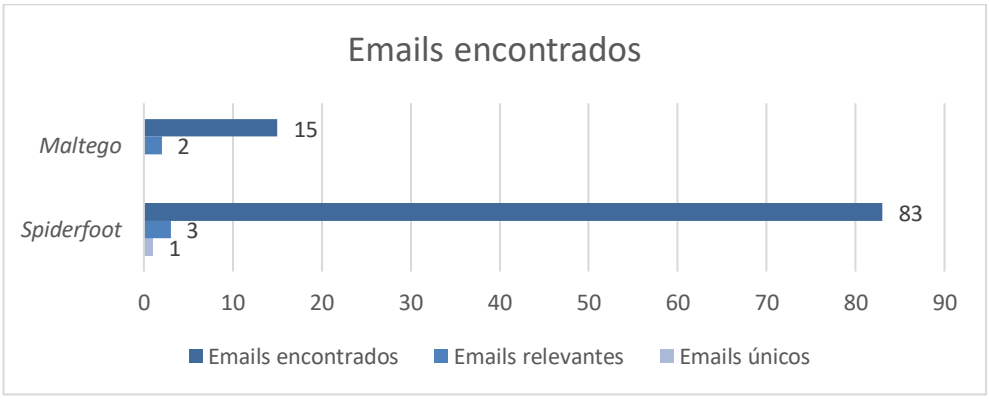


Figura 46: Emails encontrados

Según se ve en la Figura 47, con *recon-ng* se localizan pocos enlaces a redes sociales que pertenezcan al centro educativo. Con *Maltego* se localizan todas las cuentas de redes sociales

que están enlazadas en la sección principal de la *web*. Con *Spiderfoot* están los datos, pero según vemos en la Figura 48 es más difícil encontrarlos, la información está dispersa en varias categorías del informe lo que dificulta su análisis e incrementa el tiempo de extracción de información fiable. Claramente con *Maltego* se obtienen mejores resultados.

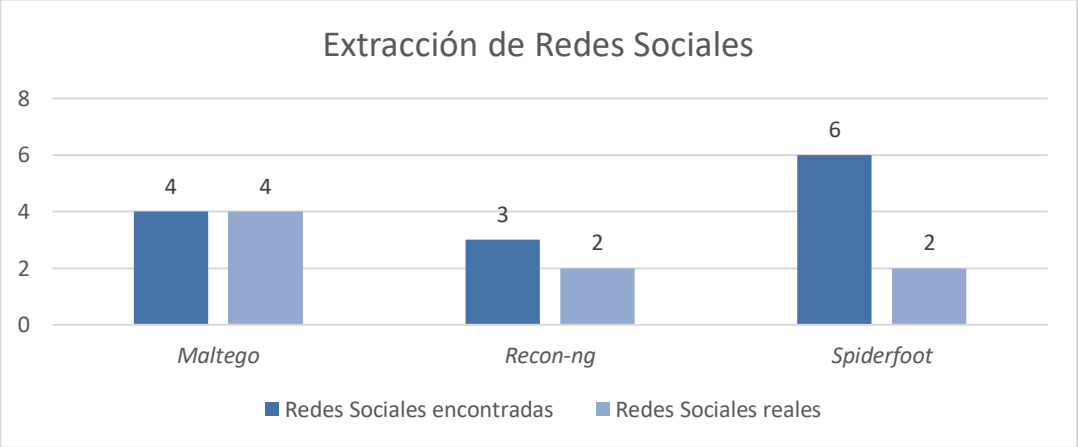


Figura 47: Extracción de Redes Sociales

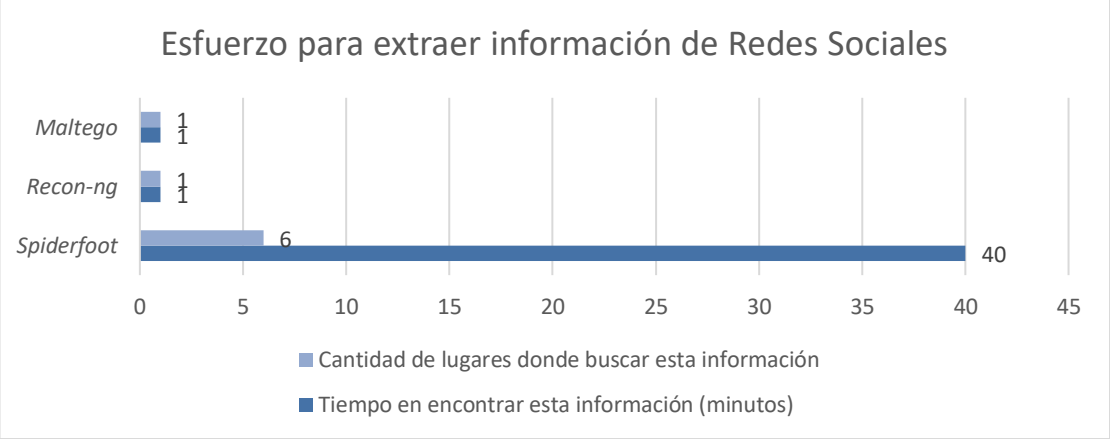


Figura 48: Esfuerzo para extraer información de Redes Sociales en los informes de las herramientas

Tras el análisis con *The Harvester*, Figura 49. Se encuentran usuarios *Linkedin* y a simple vista la mayoría parecen que son datos irrelevantes, se encuentra un solo contacto que hace referencia directa con el objetivo analizado. Se localizan varias personas que en principio no lo relacionan con el Centro Educativo, hay que analizar detalladamente estos perfiles por si fueran antiguos alumnos, personal del centro o proveedores. Puede resultar interesante si se desea profundizar por esta vía de investigación.

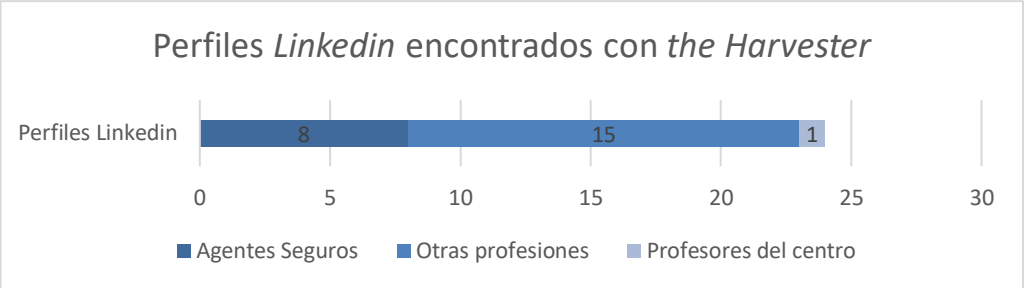


Figura 49: Perfiles Linkedin

Respecto a enlaces a otras *webs*, con *Maltego* se hace una buena recopilación de lo que hay publicado en la *web*, que comprende las páginas de la Junta de Andalucía (Moodle, Séneca, PASEN, Secretaría virtual), dos *blogs* del centro y muchos enlaces a documentos en *docs.google*. Con *Spiderfoot* se encuentra lo mismo que con *Maltego* pero además se encuentran más *blogs* de asignaturas, un enlace adicional a una *web* del centro alojado en un blog de la Junta y más enlaces a documentos en *docs.google*. Como se observa en la Figura 50 la cantidad de *URL* que se analiza es muy grande, esto requiere tiempo y esfuerzo para su análisis. *Spiderfoot* ordena los datos de diferentes maneras, por tipo, por familia, por origen. En este caso se ordena por tipo, uno de estos tipos se llama *Linked URL- External,* el cual hace referencia a 1.580 enlaces. Estos enlaces, la inmensa mayoría de ellos se repiten. Por ejemplo, un *blog* de matemáticas aparece 75 veces, lo único que cambia es que apunta a diferentes subcarpetas o directorios dentro del mismo dominio. Así sucede con muchos de los *blogs* y enlaces de la propia *web* principal.

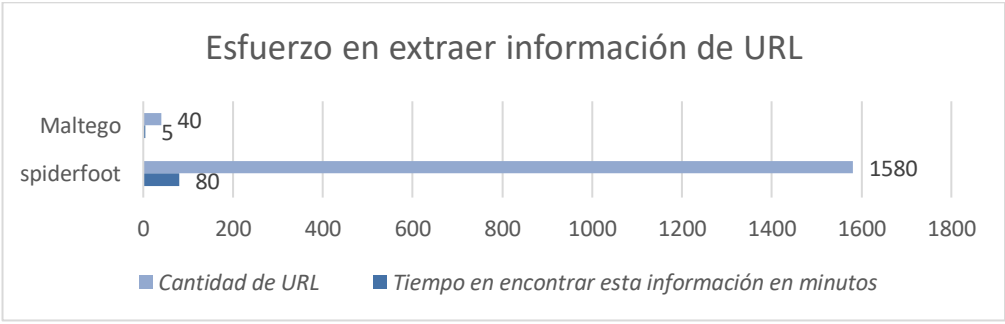


Figura 50: Esfuerzo en extraer información de URL los informes de las herramientas

Entre los enlaces repetidos y los enlaces no relevantes, al final queda poca información importante respecto a la cantidad que se partía inicialmente, ver la Figura 51. Sigue siendo más información relevante la que consigue respecto *Maltego* pero penaliza el tiempo que hay

que emplear para extraer esta información. Como se observa se hace una representación, con cada herramienta, de la cantidad de estos enlaces por tipos: de *Google Doc*; de la Junta de Andalucía y de *blogs.*

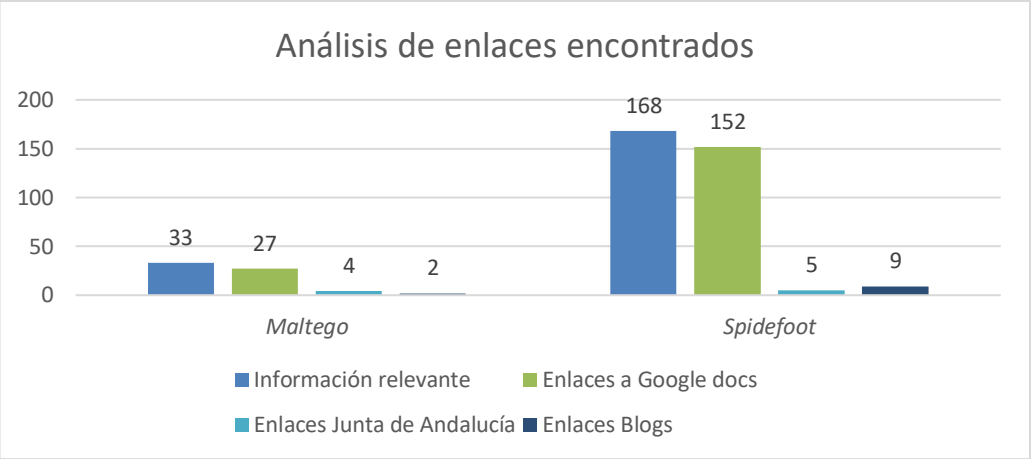


Figura 51: Análisis de enlaces encontrados

6.1.3 Software, puertos, IP y nombres reales

La columna del software utilizado de la Tabla 12, se refiere a los programas que han podido utilizar para la generación de contenido en la página *web*. Información que se encuentra entre otros por del análisis de los metadatos.

Tabla 12: Software, puertos, IP y nombres reales

	Software utilizado	Puertos	IP compartida	Nombres Reales
Spiderfoot	SI	Si	Si	Si
Foca	No	Si	No	No
Shodan	No	Si	No	No
Repositorios	No	No	Si	No

En este caso se esperaba encontrar información con *FOCA,* pero no se encontró nada. Sin embargo, con *Spiderfoot* se encuentra mucha información de software utilizado, incluso detectando algún sistema operativo.

Como se observa en la Figura 52, con *FOCA* se encuentra un grupo reducido de puertos TCP. Con *Shodan* y *Spiderfoot* se encuentran más puertos abiertos TCP e incluso un puerto UDP. Hay puertos que solo se detectan con *Spiderfoot* y otros puertos que solo han sido detectados con *Shodan*. De un total de quince puertos abiertos detectados, sumando todas las herramientas. Con *Spiderfoot* se localizan cuatro puertos que no se encuentran con las otras herramientas y con *Shodan* son dos puertos. En todas las herramientas este dato es fácilmente localizable.

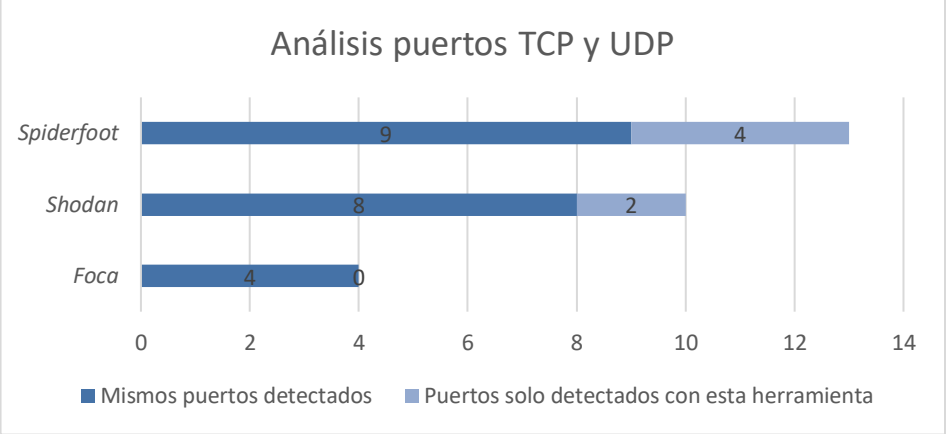


Figura 52: Análisis puertos TCP/UDP encontrados

Respecto a la columna de IP compartida de la Tabla 12, hace referencia a otros dominios con la misma IP que nuestro objetivo. Indica que comparten *hosting* y es una información que puede ser importante. Con una de las herramientas del repositorio (*builtwith.com*) se consigue esta información, pero también con *Spiderfoot* y de manera automática.

Por último, los nombres reales, son nombres de personas que detecta la aplicación. Nombres que aparecen en el dominio. Según vemos en la Figura 53 *Spiderfoot* localiza muchos de ellos, y otros muchos nombres que no se corresponden con nombres de reales de personas, pero a partir de estos nombres extiende su búsqueda a otro tipo datos, como pueden ser blogs, redes sociales o cuentas en otros sitios.

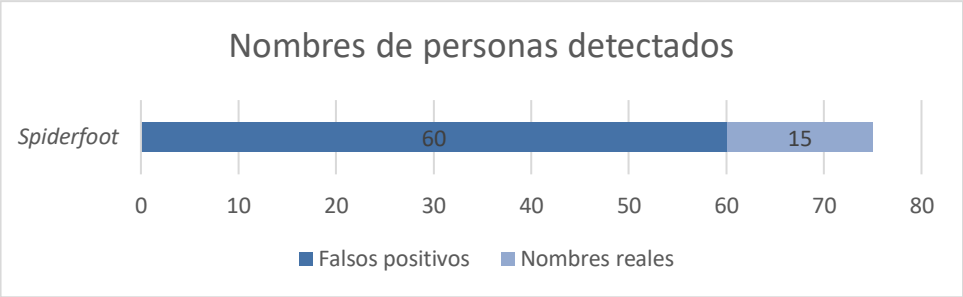


Figura 53: Nombres detectados

6.1.4 Enlaces documentos Google

Maltego y *Spiderfoot* encuentran más de 140 enlaces a documentos en Google, estos enlaces son públicos. Con *Maltego* no funcionan los enlaces porque no ha respetado las mayúsculas y minúsculas a la hora de recopilarlos. Sin embargo, con *Spiderfoot* sí que se han respetado y funcionan los enlaces. De todas maneras, toda esta información se encuentra en la *web* de centro.

Debido a la pandemia y la necesidad de seguir prestando la formación a distancia al alumnado, los centros han recurrido al uso de herramientas de todo tipo, aparte de las proporcionadas por las autoridades competentes. En este caso se ven en estos documentos que el centro ha recurrido *Google Classroom*, *Google Drive*, correo electrónico y redes sociales. La información que se ha encontrado en estos documentos es la siguiente:

- Claves de *Google Classroom* de todas las asignaturas y cursos.
- Correo electrónico de personal docente. Correos personales, correos para este propósito y correos corporativos.
- Redes sociales. *Youtube, Telegram* o *Instagram*. En este último incluso alguno público en el que se ve el perfil de los seguidores, en su mayoría alumnos. Figura 54.
- Enlaces a Google Drive. En uno de estos enlaces se ha podido encontrar nombre completo y fotografía de una clase. Figura 55.

Figura 54: Enlaces a Instagram

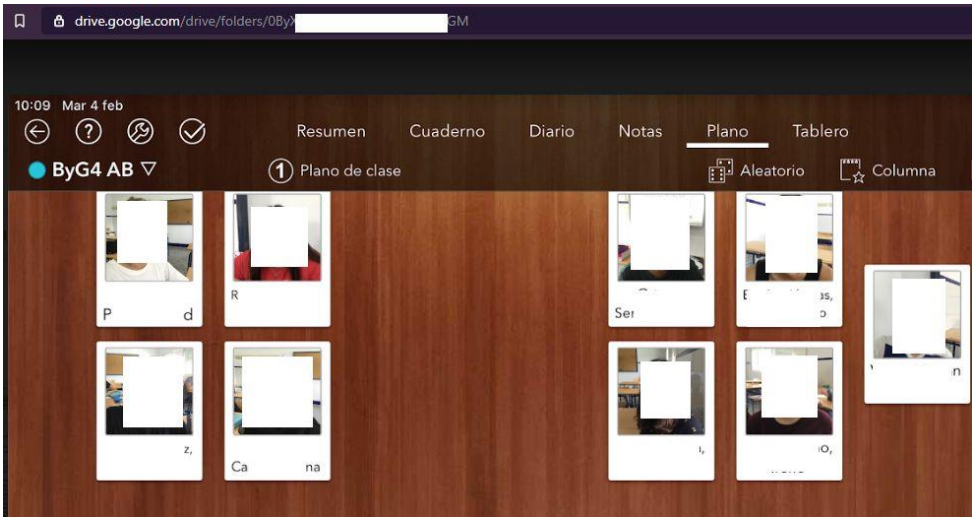


Figura 55: Fotografía, nombre completo y curso académico

En la Figura 56 se puede observar parte de una captura de uno de estos documentos compartidos, en este caso se aprecia el código *Classroom* de la asignatura y curso concreto, además de una cuenta de correo, en este caso corporativo. Esto es así en todos los documentos, se encuentra muchos más correos electrónicos, códigos de *Google Classroom* y perfiles de redes sociales. Lo interesante en este caso en el nombre usuario del correo, se compone de siete letras seguido de tres números. Esta codificación se corresponde con el Identificador Educativo Andaluz (IdEA) y es el nombre de usuario se utiliza para acceder a cualquier aplicación de la Consejería de Educación de las que se han comentado antes: Séneca, Moodle o PASEN.

Este nombre de usuario, se puede encontrar información de cómo se compone en una búsqueda sencilla en internet, en blogs o en la propia *web* de la Consejería. Se coge la letra inicial de tu nombre, seguido de las tres primeras letras de tu primer apellido, las tres primeras

letras de tu segundo apellido y los tres últimos números de tu DNI. Es por ello que viendo el nombre del docente publicado en la *web* podemos deducir que ese nombre de usuario del correo, es su usuario IdEA.

Como se menciona en el capítulo 2 la búsqueda avanzada de Google da muy buenos resultados. Si buscamos: site:juntadeandalucia.es/educacion filetype:pdf intext:dni, encontramos listados de docentes con nombre y apellidos y su DNI. Ya podemos componer el usuario IdEA.

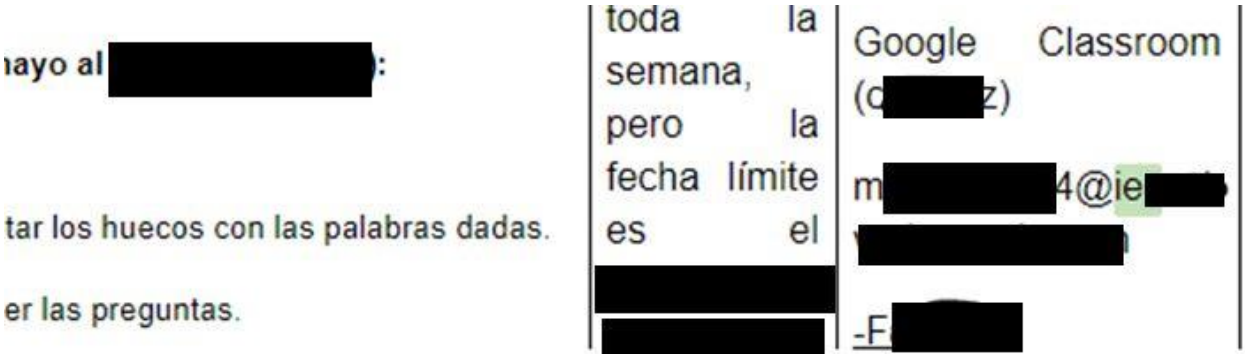


Figura 56: Códigos Classroom y cuentas de correo

6.1.5 Conclusión del análisis

Después de analizar los resultados de las herramientas estudiadas, para un objetivo tipo de un Centro Educativo de Educación Secundaria y partiendo el análisis desde la misma información, podemos extraer una serie de conclusiones.

Tras estudiar trece módulos. Con *Spiderfoot* y *Maltego* se han encontrado más información que con el resto de herramientas, ver Figura 57. También es cierto que las otras herramientas eran más específicas por lo que la comparación justa es con *Recon-ng*. Con esta última herramienta no se han obtenido los resultados esperados. Para este objetivo, partir como dato de entrada solo con el dominio, el tipo de datos que intenta encontrar y las fuentes donde busca la información, no encajan para este ámbito. Quizás hay que partir desde otro tipo de información y emplear las API específicas para realizar determinadas búsquedas. Además, esta herramienta permite crear tus propias búsquedas y orígenes de datos.

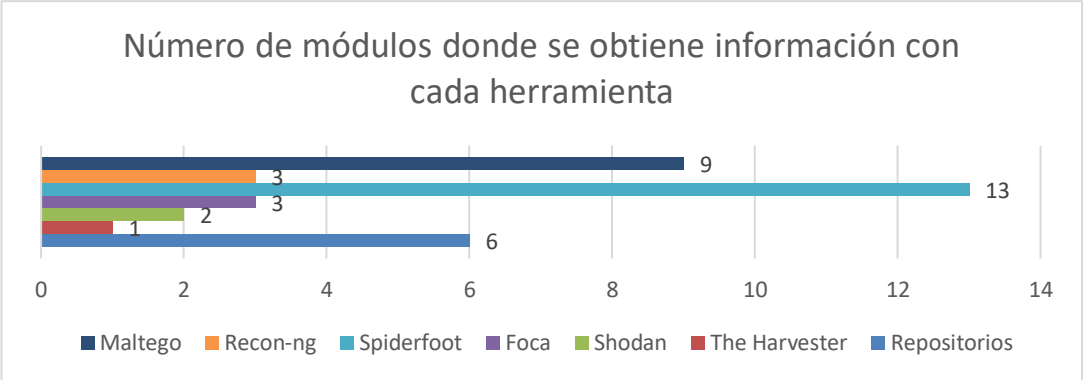


Figura 57: Módulos analizados con cada herramienta

Del repositorio de *OSINT Framework* como se explica en el capítulo anterior, se seleccionan una serie de herramientas para investigar a partir de información del dominio dado. Con varias de ellas se encuentra buena información que se complementan entre ellas, pero podemos sacar esta información de manera automática con *Maltego, Spiderfoot* y también con *FOCA*. Es verdad que también se pueden usar estos repositorios como validación de la información encontrada de manera automática.

Con *FOCA* se esperaba encontrar información del análisis de los metadatos, pero para este objetivo en concreto no se han conseguido buenos resultados. Lo que sí es interesante, es la información que proporciona de manera resumida, información del dominio, dominio del *hosting*, tecnología del servidor e incluso algún puerto y servicio. Proporciona estos resultados de forma rápida y presenta los resultados de manera ordenada. Para este contexto resulta interesante tener la información de esta manera.

Con la aplicación *Shodan* se han reportado todos los puertos publicados conociendo solo la IP del dominio, ha dado información interesante, de puerto de bases de datos, de servicios FTP, SMTP, HTTPS o DNS.

En general, para el tipo de objetivo analizado, podemos destacar por encima del resto las herramientas *Spidefoot* y *Maltego*. Con la segunda herramienta el informe final presenta los resultados de manera muy estructurada y clara. Esto permite ver de forma rápida los datos más relevantes que se pueden extraer del dominio objetivo. No presenta excesiva información que no es útil (ver Figura 58) y no se tarda mucho tiempo en obtener resultados (Ver Figura 44). Todo esto facilita su compresión y para este contexto de análisis resulta ventajoso. Pero también es cierto que la información queda muy limitada al número de transformadas que permite hacer, aunque tiene la posibilidad de crear las tuyas específicas.

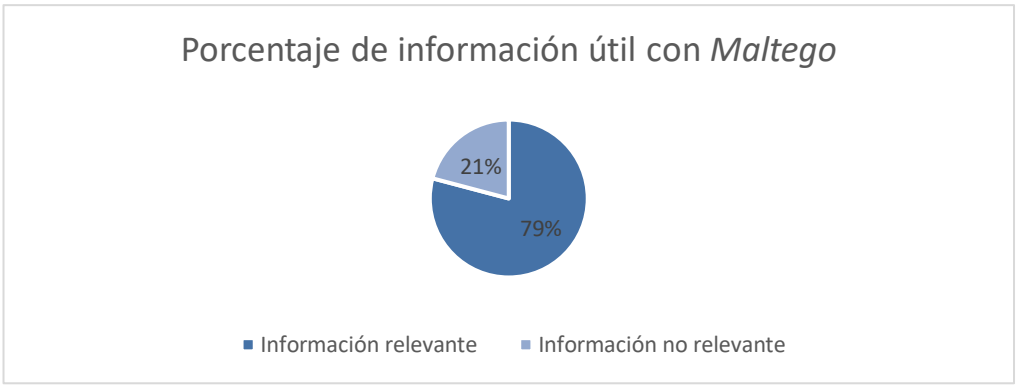


Figura 58: Porcentaje de información útil con Maltego

Con *Spidefoot* es con la que más volumen de información se extrae, alrededor de 12.000 elementos como se ve en la Figura 28. Para este caso la herramienta emplea cuatro horas en la tarea de recopilación de información (Ver Figura 44). Hay mucha información y como consecuencia mucha información irrelevante, como se puede observar en la Figura 59. Esto requiere de un análisis para descartar con rapidez información que no resulta útil. La herramienta facilita diferentes vistas para ordenar los datos, pero aun así se necesita tiempo para discernir esta información.

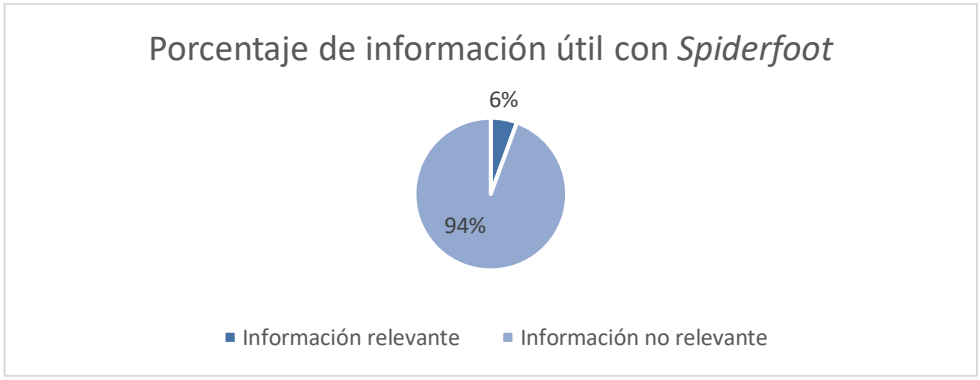


Figura 59: Porcentaje de información útil con Spiderfoot

No es un problema grande, si tienes claro cuál es el objetivo, pero hay que dedicar más de tiempo en localizar la información importante de la que no es relevante o directamente no es veraz. Esta herramienta emplea búsquedas en la *Darknet* a partir de nombres de personas que encuentra o de nombres de usuarios. Realiza búsquedas de cuentas en *blogs* y redes sociales partiendo de estos nombres. Quizás en principio, esto no es interesante en el campo de los Centros Educativos.

Si hacemos una criba de todo este mar de datos, o si afinamos más el escaneo, se localiza más información que el resto de herramientas: con esta herramienta se ha recopilado más información y datos que con el resto juntas, a pesar de que solo el 6% es información relevante. También *Spiderfoot* marca cierta información que considera que tienen más riesgo, por vulnerabilidades o tipo de información expuesta. Permite exportar los datos en formato CSV para tu propia explotación, a parte de las presentaciones gráficas mencionadas en el capítulo anterior.

En la Figura 60, se analiza la cantidad de información única que se obtiene con todas las herramientas. De los trece módulos estudiados, en siete de ellos se han encontrado información que solo se encuentra por una única herramienta. Son sesenta entidades o información en total que se obtienen de estos módulos. De los cuales treinta y seis es el número de información única que se detecta con *Spiderfoot* y dos con *Maltego* y *Shodan* cada uno. Es decir, de cuarenta datos únicos, la inmensa mayoría de ellos se obtienen con *Spiderfoot*. Pero también como se veía en la figura anterior muchos de los datos que entrega esta herramienta no resulta interesante.

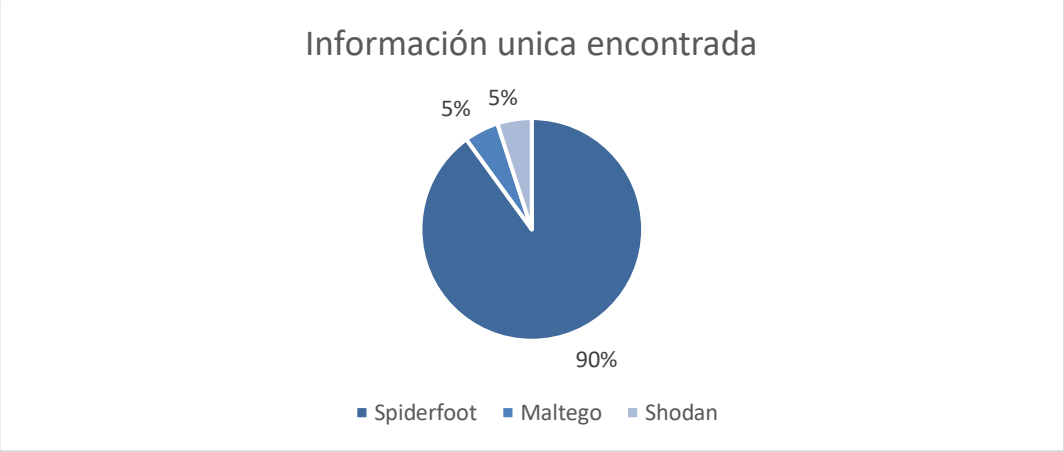


Figura 60: Porcentaje de información única detectada por una herramienta

Lugares donde deben centrarse los Centros Educativos

Para analizar una institución como la de nuestro objetivo, el resultado con *Spiderfoot* ha sido positivo, quizás para no extender demasiado la recopilación se podría afinar más el escaneo. Limitado y proponiendo otros valores de entrada o empleando la opción alternativa al escaneo que es el que denomina *investigación*, que vas poco a poco extendiendo la recopilación de datos a medida que vas encontrando información, un análisis más parecido a lo que puedes hacer con *Maltego*. Antes o después de realizar un análisis largo de este tipo, con *Maltego* se

puede realizar una recopilación bastante más rápida, pero eso sí, más limitada pero que puede ayudar a no perderse en toda la información recopilada con *Spiderfoot* y centrarte en lo importante. Sirve de buen complemento para realizar una búsqueda de calidad de información y de criba de datos. Y por último con *FOCA* y *Shodan,* aunque puede ser repetitivo, se obtiene una información muy concreta, rápida y bien resumida que corrobora los datos encontrados e incluso los amplia en algún caso.

Por lo tanto, para que los centros educativos saquen el máximo rendimiento a estas herramientas, se propone centrarse en ciertos aspectos en cada una de ellas:

Maltego

- Centrarse en todo el abanico de transformadas denominadas *To DNS.* Con los que se puede obtener información valiosa sobre el dominio, subdominios, registros NS, registros MX.
- Datos que se obtienen con la transformada *Mirror: External Links*. Es información valiosa, de enlaces a otras páginas relevantes, a blogs y redes sociales. Como se ha reseñado, quizás su punto fuerte sea el apartado de redes sociales.

Spiderfoot

- Para la búsqueda de subdominios centrarse en las fuentes *DNS-Brute-Force* y *DNS-Grep*. Además del apartado *Internet Name* que recopila toda esta información.
- Centrarse en el apartado de búsqueda de Emails, en las fuentes *Email-Extractor*, *Email Address* y *Affiliate Email Address*.
- Para localizar enlaces a otras *Web*s, como son los blogs. Analizar el apartado *Linked URL- External.*
- Para analizar el Software que utiliza el Centro centrarse en el apartado *Software Used*.
- Para el análisis de puertos centrarse en los apartados *Open TCP Port*, *Open UDP Port* y *Open TCP Port Banner*.
- Para ver el hosting compartido analizar los apartados *Co- Hosted Site*.
- Los nombres reales detectados, localizados en *Human Name.*
- Para la recopilación de tecnología utilizada en los servidores, analizar los datos de *Web Server, Web Technology* y *Web Framework.*

FOCA

- Centrarse en la información relevante a la tecnología y datos del dominio. Informe muy claro y estructurado.

Shodan

- Los puertos interesantes que se pueden detectar de manera muy rápida. Y si se dispone de más información de otras IP, es mucho más provechoso.

6.2 Propuesta de buenas prácticas

Como hemos podido comprobar, el uso de estas herramientas nos ayuda a detectar información y datos que estamos exponiendo de los centros educativos y no se es consciente de ello. Por el tipo de organización de la que se está analizando, en el que hay cierta libertad de toma de decisiones y la falta de directrices claras por parte de las instituciones. Por la escasez de recursos, en lo que no todos los procedimientos están regulados. Esto hace que se cometan ciertos errores estructurales, que ponen en riesgo y comprometen los datos personales de docentes y alumnos, que además en su mayoría son de especial protección por ser menores de edad.

El uso de estas herramientas, como parte de una solución global es buena para empezar a auditar y detectar estas deficiencias. A continuación, enumeramos una breve propuesta de buenas prácticas para proteger mejor su información en el ámbito educativo.

- Elaboración de una Política de Seguridad. Que sea clara y, lo más importante, asegurarse de que todo el mundo la cumpla. Tanto internamente como los proveedores externos de servicios. Por ejemplo, la norma ISO 27001:2017 en el apartado 5.2 puede servir como guía.

- Elaborar un documento de seguridad que recoja medidas técnicas y organizativas referentes a la normativa vigente en seguridad. Uso responsable y seguro de las TIC, inculcar al alumnado y docentes comportamientos adecuados en el uso de las redes sociales o en general en el uso de los equipos informáticos. Que abarque desde las redes internas, tanto cableada como inalámbrica, dispositivos externos e internos, ordenadores, uso de la página *web*, redes sociales, correo electrónico y uso de las aplicaciones.

- Tener a disponible un procedimiento de notificación de incidentes de seguridad o incidencias, en la que el personal y alumnado puedan notificar los posibles errores, incidencias o fugas de datos detectados.

- Realizar auditorías internas y externas. Esto permite ver el estado actual de tu sistema, los datos expuestos, los nuevos riegos y ver que deficiencias a mejorar.

- Definir y documentar las personas con accesos a los datos para determinar sus funciones y obligaciones. Importante para saber en todo momento quien es el responsable de los datos y que se cumplan las directrices de seguridad. Por ejemplo, las personas que tienen acceso al correo electrónico del centro.

- Crear un listado de las personas con acceso a información de carácter personal, de los responsables de seguridad o de cualquier figura de este tipo. En los centros educativos no hay una separación clara de funciones y estas normativas ayudan a proteger mejor los datos.

- Políticas de usuarios y contraseñas. Cada empleado o docente su usuario. Las contraseñas no compartirlas, que sean diferentes para cada servicio, y definir complejidad y caducidad. Importante de cara a las aplicaciones tipo *Seneca* de los docentes, o la tendencia de usar la misma sesión en los equipos informáticos los docentes o administrativos, y mucho menos mezclar usuarios y contraseñas de uso personal con el profesional.

- Definir una gestión de mínimos privilegios. De esta manera evitamos que por error o de manera malintencionada personas no autorizadas puedan acceder o modificar datos que no deberían ver.

- Copias de seguridad. Básico para cualquier sistema de información. Por ejemplo, la página *web* puede ser víctima de un ataque al explotar alguna vulnerabilidad detectada. Se deben disponer de copias.

- Gestión de soportes y documentación. Registro de acceso a la información. Importante que exista un seguimiento de la información, para poder determinar una posible brecha o fuga de datos.

- Cifrado de dispositivos de almacenamiento. *Pendrives*, discos duros internos y externos. Estos dispositivos son susceptibles de perderse o dejarlos olvidados, en el

que pueden contener datos confidenciales del personal o alumnado, por costumbre debe estar cifrado con contraseña.

- Capacitación en seguridad de redes escolares para el personal y los estudiantes. Explicar claramente el buen funcionamiento y uso. Conocimientos básicos de uso de antivirus, la existencia del malware y navegación privada.

- Formación del uso correcto de internet a docentes, personal administrativo y alumnado. Entender la importancia de la privacidad en la red, que tengan conocimiento de cuáles son las actividades ilícitas en la red y entiendan conceptos de sitios seguros e inseguros. Identificar el *phishing* u otras estafas.

- Los riesgos en las redes sociales. Protección de datos de carácter personal e intelectual. Protección de la privacidad, honor, intimidad y propia imagen. Información y colaboración con las familias. Esto puede evitar que los docentes, por ejemplo, publiquen en un *Google Drive* público las imágenes y nombres de los alumnos. También que cree un perfil en una red social que la configuración sea pública, donde se vean todos los seguidores, que en realidad son sus alumnos.

- No difundir en las redes sociales y por ende Internet, ningún tipo de imagen, foto o video de cualquier miembro de la comunidad educativa: Los alumnos y alumnas del centro de enseñanza, deberán evitar subir imágenes propias e información que los pueda identificar a las redes sociales sin el permiso de la familia.

- No publicar contraseñas de ningún tipo, ni perfiles de redes sociales. Transmitir esta información por los mecanismos dotados por las Instituciones por ejemplo en este caso estudiado, por *PASEN*. Tener contraseñas de *Classroom* publicados y de perfiles de redes sociales o correos electrónicos es un riesgo enorme.

7. Conclusiones y trabajo futuro

Los Centros Educativos son un objetivo de los ciberataques por el tipo de información que poseen. La digitalización en la que están inmersos y la educación a distancia motivada por la pandemia mundial, ha provocado que las nuevas amenazas y riesgos crezcan aún más y que haya mucha más información expuesta. Es por ello que, estudiar las herramientas OSINT para determinar cuáles de ellas pueden dar mejores resultados en este contexto, supone dotar a estos Centros de mecanismos para conocer sus riegos, realizar mejor la prevención de fuga de datos y limitar la información expuesta. En definitiva, ser conscientes de los errores que se cometen con el manejo de la información, las redes, y el ciberespacio en general y así estas instituciones ser capaces de protegerse mejor. A continuación, se incluye un resumen de las contribuciones del trabajo:

- A través del estudio en profundidad del estado del arte, se conoce OSINT, como es su funcionamiento, la metodología empleada en estas técnicas y sobre todo los retos y las ventajas que supone utilizar este método en diferentes ámbitos y en el educativo en particular.
- Se presentan las tendencias y estudios relevantes que usan esta técnica, que dan una visión general de los campos de aplicación y potencial que tiene el uso de OSINT en prácticamente todos los sectores.
- Conocer cuál es el nivel de digitalización y estado actual de las instituciones educativas ha permitido comprender que este sector está inmerso en una revolución digital y la tendencia es ascendente. El sector educativo es un objetivo sustancial en los ciberataques y que en muchos casos en esta digitalización no se están teniendo en cuenta muchos factores por las prisas y falta de rigor.
- Se han estudiado las herramientas más relevantes de OSINT, para comprender el tipo de información que se puede obtener con ellas en el sector educativo. Herramientas más generales (*Maltego, Recon-nf y Spiderfoot*) que recopilan información de todo tipo, y también herramientas especificas (*FOCA, Shodan, The Harvester y Repositorios*) orientadas a un tipo de dato.
- Se elabora un experimento en el cual se hace uso de estas herramientas, aplicándolas a un Centro Educativo de Educación Secundaria. Se elige un Centro Educativo en España, que puede ser uno estándar a nivel nacional. Y se configuran todas las herramientas de manera que todas parten de la misma información.
- Tras el análisis de los resultados obtenidos con todas las herramientas. Se hace una valoración de los puntos relevantes y se concluye qué herramientas funcionan mejor

para este tipo de institución. El uso de estas herramientas es relativamente sencillo, además son herramientas de uso libre, se adaptan perfectamente a un entorno educativo por el tipo de datos expuestos, por lo general no presentan problemas legales y además permiten añadir otros orígenes de datos. Hacen que la solución planteada encaje perfectamente con un Centro Educativo donde no existe un perfil técnico.

- Se hace una reseña de donde centrar el análisis en estas herramientas, siempre en nuestro contexto. Es decir, que resultados de los que nos devuelven las herramientas, hay que tener más en consideración a la hora de extraer conclusiones.
- A partir de las deficiencias encontradas se elabora una guía de buenas prácticas, para que los centros educativos tengan una serie de puntos de mínimos a considerar, para proteger mejor su información expuesta.

Como trabajos futuros se plantean los siguientes puntos:

- Las Administraciones públicas pueden emplear esta herramienta, sin incurrir en demasiados costes y conocer el estado de sus Centros Educativos, por ejemplo, realizando auditorías al azar. Con el objetivo de elaborar un plan de acción de mejora y coordinar mejor su cumplimiento.
- La administración proponga una metodología con la que auditar sus centros de forma anual.
- Añadir transformadas específicas en la herramienta *Maltego*. Aplicadas al contexto del ámbito educativo. Por ejemplo, lo que se conoce por Área de Influencia. En la *web* de la Junta de Andalucía, según en la calle que vives puedes ver los puntos que te corresponden para un posterior baremo y así acceder a un centro educativo determinado por la distancia a tu vivienda. O también, automatizar las búsquedas del personal docente en la *Web* de la Consejería de Educación o el Ministerio.
- Elaborar un análisis paso a paso con *Spiderfoot* para limitar el número de información irrelevante y, de esta manera dirigir la búsqueda de información hacia los sitios que se han obtenido mejores resultados. Con el fin de crear perfiles específicos y automatizar este proceso.
- Anadir tanto a *Maltego* como a *Spiderfoot,* las *Key* de las *API* que son requisito necesario para algunas búsquedas.

8. Bibliografía

Alves, F., & Ferreira, P. M. (2018). *OSINT-based Data-driven Cybersecurity Discovery*. 4.

Bellincat. (s. f.). bellingcat. Recuperado 16 de abril de 2020, de https://www.bellingcat.com/

Cartagena, A., Rimmer, G., van Dalsen, T., Watkins, L., Robinson, W. H., & Rubin, A. (2020). Privacy Violating Opensource Intelligence Threat Evaluation Framework: A Security Assessment Framework For Critical Infrastructure Owners. *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 0494-0499. https://doi.org/10.1109/CCWC47524.2020.9031172

Cerny, J., Potančok, M., & Molnar, Z. (2019). Using open data and Google search data for competitive intelligence analysis. *Journal of Intelligence Studies in Business*, *9*, 72-81. https://doi.org/10.37380/jisib.v9i2.470

Criado, J. M. (2020, enero 24). *Estas son las 6 principales ciberamenazas de 2020*. Terránea. https://blog.terranea.es/ciberamenazas-2020/

Dossier de Indicadores de la Sociedad Digital por género (marzo 2020) | *Ontsi—Red.es*. (s. f.). Recuperado 11 de mayo de 2020, de https://www.ontsi.red.es/es/dossier-de-indicadores-pdf/Dossier-de-Indicadores-de-la-Sociedad-Digital-por-genero-%28marzo-2020%29

Dossier de Indicadores relacionados con la administración electrónica y las TIC en la educación en España (febrero 2019) | *Ontsi—Red.es*. (s. f.). Recuperado 11 de mayo de 2020, de https://www.ontsi.red.es/dossier-de-indicadores-pdf/Dossier-de-Indicadores-relacionados-con-la-administracion-electronica-3

EE_La_transformacion_digital_del_sector_educacion-1.pdf. (s. f.). Recuperado 15 de abril de 2020, de http://www.fundacionorange.es/wp-content/uploads/2016/11/eE_La_transformacion_digital_del_sector_educacion-1.pdf

El sector educativo, un blanco perfecto para ciberataques. (s. f.). Recuperado 11 de mayo de 2020, de https://www.marsh.com/pa/es/insights/risk-in-context/sector-educativo-ciberataques.html

E-Learning Market Trends 2020-2026 | Global Research Report. (s. f.). Global Market Insights, Inc. Recuperado 10 de mayo de 2020, de https://www.gminsights.com/industry-analysis/elearning-market-size

Estadísticas de la Educación. (s. f.). Recuperado 11 de mayo de 2020, de https://www.educacionyfp.gob.es/servicios-al-ciudadano/estadisticas.html

European Monitoring Centre for Drugs and Drug Addiction. (2019). *Using open-source information to improve the European drug monitoring system*. 33.

Fantinelli, S., & Sivilli, D. F. (2015). Open Source Intelligence's Methodology Applied to Organizational Communication. *Mediterranean Journal of Social Sciences*, *6*(2), 233.

Fleisher, C. (2008). Using Open Source Data in Developing Competitive and Market Intelligence. *European Journal of Marketing*, *42*, 852-866. https://doi.org/10.1108/03090560810877196

FOCA. (s. f.). Recuperado 16 de abril de 2020, de https://www.elevenpaths.com/es/labstools/foca-2/index.html

Fogelman-Soulié, F. (2008). *Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security*. IOS Press.

Hassan, N. A., & Hijazi, R. (2018). *Open Source Intelligence Methods and Tools: A Practical Guide to Online Intelligence* (1st ed.). Apress.

Hernandez Mediná, M. J., Pinzón Hernández, C. C., Díaz López, D. O., Garcia Ruiz, J. C., & Pinto Rico, R. A. (2018). Open source intelligence (OSINT) in a colombian context and sentiment analysis. *Revista Vínculos*, *15*(2), 195-214. https://doi.org/10.14483/2322939X.13504

https://plus.google.com/+UNESCO. (2020, marzo 4). *COVID-19 Educational Disruption and Response*. UNESCO. https://en.unesco.org/covid19/educationresponse

Johnson, L. K. (2007). *Handbook of Intelligence Studies*. Routledge.

Las ciberamenazas ponen en alerta a las universidades. (s. f.). Deloitte Spain. Recuperado 11 de mayo de 2020, de https://www2.deloitte.com/es/es/pages/governance-risk-and-compliance/articles/ciberamenazas-alertan-universidades.html

Lu, J. (2020). *User churning behavior in social networks*. http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-48014

Manual de verificación: Página de inicio. (s. f.). Recuperado 7 de mayo de 2020, de http://verificationhandbook.com/book/

Marco OSINT. (s. f.). Recuperado 5 de mayo de 2020, de https://osintframework.com/

Martín-Arroyo, J. (2020, abril 29). *Millones de datos de alumnos y profesores están expuestos por la educación 'online'*. EL PAÍS. https://elpais.com/sociedad/2020-04-29/millones-de-datos-de-alumnos-y-profesores-estan-expuestos-por-la-educacion-online.html

Mercado MOOC por Plataformas y Servicios—2023 | Mercados y Mercados. (s. f.). Recuperado 11 de mayo de 2020, de https://www.marketsandmarkets.com/Market-Reports/massive-open-online-course-market-237288995.html

Miniwatts Marketing Group. (2020). *World Internet Users Statistics and 2020 World Population Stats*. https://www.internetworldstats.com/stats.htm

MOOC Market Forecast, Trend Analysis & Competition Tracking—Global Market Insights 2019 to 2029. (s. f.). Fact.MR. Recuperado 11 de mayo de 2020, de https://www.factmr.com/report/3077/mooc-market

Nato OSINT Handbook v1.2—Jan 2002. (2001). 57.

OSINT - La información es poder. (2014, mayo 28). INCIBE-CERT. https://www.incibe-cert.es/blog/osint-la-informacion-es-poder

Pastor-Galindo, J., Nespoli, P., Gómez Mármol, F., & Martínez Pérez, G. (2020). The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends. *IEEE Access*, *8*, 10282-10304. https://doi.org/10.1109/ACCESS.2020.2965257

Ph055a. (2020). *Ph055a/OSINT_Collection*. https://github.com/Ph055a/OSINT_Collection (Original work published 2018)

Plan Digital 2020: La digitalización de la sociedad española. (s. f.). 132.

Portillo, I. (2019, octubre 17). *Descubriendo que son los Data Leaks y los Data Breaches*. Derecho de la Red. https://derechodelared.com/data-leaks-data-breaches/

Saballs, J. T. (2005). *Los centros educativos como organizaciones*. 61.

SecurityTrails | Top 25 OSINT Tools for Penetration Testing. (s. f.). Recuperado 3 de julio de 2020, de https://securitytrails.com/blog/osint-tools

The Digital Vortex in 2019: Continuous and Connected Change. (s. f.). IMD business school. Recuperado 10 de mayo de 2020, de /research-knowledge/reports/digitalvortex2019/

Verification Handbook. (s. f.). DataJournalism.com. Recuperado 7 de mayo de 2020, de https://datajournalism.com/read/handbook/verification-3

Williams, H., & Blum, I. (2018). *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*. RAND Corporation. https://doi.org/10.7249/RR1964